

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:05:10 : Search time 19.4298 Seconds  
(without alignments)  
1995.696 Million cell updates/sec

Title: US-09-490-291-2  
Perfect score: 1485  
Sequence: 1 MRGSHHHHGHGSMASGRGGL.....YGGLSQGTSGIRPRAKLN 291

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1485	100.0	291	22	AAB82608 Spider recombinant
2	1425	96.0	681	22	AAB82609 Spider recombinant
3	1357	91.4	691	22	AAB82610 Spider recombinant
4	1037	69.8	989	23	AAM50038 N. clavipes spider
5	1036	69.8	606	16	AAR99053 Spider dragline va
6	1036	69.8	606	20	AAV40100 Polymer of an anal
7	1031.5	69.5	630	23	AAM50042 N. clavipes spider
8	1031.5	69.5	676	23	AAM50047 N. clavipes spider
9	1031.5	69.5	1255	23	AAM50037 N. clavipes spider
10	1031.5	69.5	1880	23	AAM50039 N. clavipes spider

11	1031	69.4	646	18	AAW27178 Nephila clavipes s
12	1015	68.4	651	20	AAV40097 Spider silk protei
13	1015	68.4	651	23	AAU11781 Spider natural sil
14	1015	68.4	718	19	AAV53346 Nephila clavipes s
15	1015	68.4	718	21	AAV59070 N. clavipes spider
16	1008	67.9	528	22	AAB82611 Spider recombinant
17	1001	67.4	718	12	AAV14308 N.clavipes draglin
18	984	66.3	604	16	AAR99057 Spider dragline va
19	977	65.8	364	23	AAM50043 N. clavipes spider
20	977	65.8	773	23	AAM50045 N. clavipes spider
21	977	65.8	777	23	AAM50046 N. clavipes spider
22	976	65.7	606	16	AAR99055 Spider dragline va
23	976	65.7	606	20	AAV40101 Polymer of an anal
24	976	65.7	606	20	AAV40102 Polymer of an anal
25	976	65.7	809	23	AAU11793 Dragline protein 1
26	976	65.7	818	23	AAU11797 Dragline protein 1
27	976	65.7	1617	23	AAU11794 Dragline protein 1
28	976	65.7	1626	23	AAU11798 Dragline protein 1
29	922	62.1	271	23	AAM50044 N. clavipes spider
30	899	60.5	655	23	ABB76672 Protein related to
31	859.5	57.9	230	23	AAM50035 N. clavipes spider
32	693	46.7	170	23	AAM50041 N. clavipes spider
33	677.5	45.6	615	20	AAV40099 Spider silk protei
34	674	45.4	831	16	AAR80168 pMIS1 Misp spider
35	610.5	41.1	641	20	AAV28843 Epstein Barr Virus
36	610.5	41.1	641	21	AAV95856 Epstein Barr virus
37	610.5	41.1	641	22	AAV62332 EBV tethering prot
38	593.5	40.0	738	19	AAV56163 New DNA sequence 1
39	591.5	39.8	261	19	AAV79137 FLGA Gly-ala inser
40	580	39.1	264	23	AAM50048 N. clavipes spider
41	579	39.0	219	23	AAM50040 N. clavipes spider
42	561	37.8	235	17	AAW05704 Glycine-rich repea
43	561	37.8	235	19	AAV79126 Epstein Barr Virus
44	554	37.3	531	20	AAV40098 Spider silk protei
45	554	37.3	595	12	AAR14309 N.clavipes draglin

## ALIGNMENTS

RESULT 1  
AAB82608  
ID AAB82608 standard; Protein; 291 AA.  
XX AAB82608;  
XX  
XX  
DT 02-OCT-2001 (first entry)  
XX  
DE Spider recombinant silk protein pQE(spl)7.  
XX  
KW Spider; orb-weaver; silk protein; pQE(spl)7; structural protein;  
KW purification; fibre; spinning.  
XX  
OS Nephila clavipes.  
FH Key Location/Qualifiers  
FT Misc-difference 232  
FT /note= "encoded by GGY"  
XX  
PN WO200153333-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 01-NOV-2000; 2000WO-US30086.  
XX  
PR 20-JAN-2000; 2000US-0490291.  
XX  
PA (MELL/) MELLO C M.  
PA (ARCI/) ARCIDIAONO S.  
PA (BUTL/) BUTLER M M.  
PA (USSA ) US SEC OF ARMY.  
XX  
PI Mello CM, Arcidiacono S, Butler MM;









```

Db 528 GYGGLGSQG-AGRGGLGGGAGAAAAAGGAGGGGGLGSQAGGAGAGAAAAAGGAGGG 586
QY 272 YGGLGSQG 279
Db 587 YGGLGSQG 594

RESULT 6
AA40100
ID AAY40100 standard; protein; 606 AA.
XX
XX AAY40100;
AC
XX 19-NOV-1999 (first entry)
DT
XX Polymer of an analogue of spider silk protein spidroine major 1.
DE
XX Spider silk protein; spidroine major 1; cosmetic; make-up;
KW dermatological compositions; hair care; skin care; sunscreen;
KW hormone; moisturizer; skin disorder; skin disorder.
XX
OS Synthetic.
OS Nephila clavipes.
XX
XX key Location/Qualifiers
FH Peptide 1..101
FT /note= "monomer unit"
FT
XX FR2774588-A1.
XX
XX 13-AUG-1999.
PD
XX 11-FEB-1998; 98FR-0001614.
XX
XX 11-FEB-1998; 98FR-0001614.
XX
XX (OREA ) L'OREAL SA.
XX
XX Philippe M, Garson JC, Arraudeau JP;
PI WPI; 1999-510729/43.
XX
XX Cosmetic or dermatological composition containing spider silk protein,
XX for hair or skin care, in make-up or sunscreens
PS Claim 8; Flg 4B; 32pp; French.
XX
XX The present sequence represents a polymer of an analogue of the spider
XX silk protein spidroine major 1. The protein improves the moisturizing/
XX softening action of the compositions. The protein, and its fragments
XX are used in cosmetic or dermatological compositions. These compositions
XX have use as hair or skin care products; and make-up or sunscreens.
XX As the protein is a good, persistent film-formers on the skin
XX of low surface density, it can be used for delivery of active
XX agents that are generally difficult to administer, e.g. vitamins,
XX hormones, moisturizers or agents for treating disorders of the
XX skin and hair.
XX
SQ Sequence 606 AA;
Query Match 69.8%; Score 1036; DB 20; Length 606;
Best Local Similarity 64.9%; Pred. No. 5.6e-73;
Matches 239; Conservative 4; Mismatches 23; Indels 102; Gaps 13;

QY 11 GSMASRGGLGGGAGAAAA-----AAAAAGGAGGGYGGGLGS 49
Db 230 GSGAGRGGLGGGAGAAAAAGGAGGGGGLGSQAGGAGAAAAAGGAGGGYGGGLGS 289
QY 50 QGT--SGRGLGGGGA---GAAAAAAGGAGGGYGGGLGSQGTSGRGLGGGAGAA 104
Db 290 QGAGGGYGGGLGSQAGRGGGGAGAAAAAGGAGGGYGGGLGSQ-AGRGGLGGGAGAA 348
QY 105 AA-----AAAAAGGAGGGYGGGLGSQGT--SGRGLGGGGA-- 139

```

```

Db 349 AAAAGAGGGGGLGSQAGGAGAAAAAGGAGGGYGGGLGSQAGGAGGGYGGGLGSQAGR 408
QY 140 -GAAAAAAGGAGGGYGGGLGSQGTSGRGLGGGAGAAAA----- 182
Db 409 GGCAGAAAAAGGAGGGYGGGLGSQ-AGRGGLGGGAGAAAAAGGAGGGYGGGLGSOGA 467
QY 183 -----AAAAAGGAGGGYGGGLGSQGT--SGRGLGGGGA---GAAAAAAGGAGGG 232
Db 468 GQAGAAAAAGGAGGGYGGGLGSQAGGGYGGGLGSQAGRGGGGAGAAAAAGGAGGG 527
QY 233 GYGGLGSQGTSGRGLGGGAGAAAA-----AAAAAGGAGGG 271
Db 528 GYGGLGSQ-AGRGGLGGGAGAAAAAGGAGGGGGLGSQAGGAGAAAAAGGAGGG 586
QY 272 YGGLGSQG 279
Db 587 YGGLGSQG 594

RESULT 7
AAM50042
ID AAM50042 standard; Protein; 630 AA.
XX
XX AAM50042;
AC
XX 18-SEP-2002 (first entry)
DT
XX N. clavipes spidroin synthetic homologue S01 protein.
DE
XX Spidroin; spider; silk; fibre; film; membrane; wound; filter; S01.
KW
XX Synthetic.
OS
XX DE10113781-A1.
XX
XX 13-DEC-2001.
XX
XX 21-MAR-2001; 2001DE-1013781.
XX
XX 09-JUN-2000; 2000DE-1028212.
XX
XX 24-OCT-2000; 2000DE-1053478.
XX
XX (PKP-) IPK INST PFLANZENGENETIK & KULTURPFLANZE.
XX
XX Scheller J, Conrad U, Grosse F, Guehrs K;
XX
XX WPI; 2002-123561/17.
XX
XX N-PSDB; ABL61043.
XX
XX New DNA encoding synthetic spider silk protein, useful e.g. for closing
XX wounds, comprises modules that encode repeating units of spidroin
XX proteins
XX
XX Claim 22; Page 48-50; 88pp; German.
XX
XX This invention describes a novel DNA sequence, encoding a synthetic
XX spider silk protein, comprising modules, each comprising a group of
XX sequentially arranged oligonucleotides, each oligonucleotide encoding
XX a repeating unit of a spidroin protein. The synthetic protein has at
XX least 84% homology with the Nephila clavipes spidroin protein and is used
XX to produce synthetic fibres, films and/or membranes, particularly: (i)
XX for medical use, especially to close wounds and/or to support or cover
XX artificial organs; (ii) as adhesion surfaces for culturing cells; and
XX (iii) as filters. The synthetic proteins are very similar to native
XX spider silk proteins; can be prepared on a large scale and can be spun to
XX fibres with excellent mechanical properties (strength and elasticity).
XX Also they retain water solubility after long-term boiling in aqueous
XX solutions and since they are also soluble in organic solvents but
XX precipitated at high salt concentration, they are easily extracted and
XX purified. The modular construction of the invention facilitates
XX incorporation of additional peptide-encoding sequences, e.g. to simplify
XX purification or modulate solubility. This sequence represents the

```

```
CC synthetic N. clavipes spidroin-1 homologue S01 described in the
CC invention.
XX
SQ Sequence 630 AA;

Query Match 69.5%; Score 1031.5; DB 23; Length 630;
Best Local Similarity 59.8%; Pred. No. 1.3e-72;
Matches 242; Conservative 6; Mismatches 12; Indels 145; Gaps 13;

QY 11 GSMASGRGLGGQAGAAAA-----AAAAAGGAGGGY 44
  || :|||||
Db 79 GSGAGRGGLGGQAGAAAAAGGAGGGYGLGSGQAGRGGGQAGAAAAAGGAGGGY 138
QY 45 GGLGSGGTSRGGLGGQAGAAAAA-----AGGAGGGYGLGSGGTSRG 95
  ||||| :|||||
Db 139 GGLGSGG-AGRGGLGGQAGAAAAAGGAGGGYGLGGQAGGGYGLGSGG-AGRG 196
QY 96 LGGQAGAAAA-----AAAAAGGAGGGYGLGSGGTSRGGLG 135
  |||||
Db 197 LGGQAGAAAAAGGAGGGYGLGGQAGGGYGLGSGG-----AGRG 252
QY 136 GQAGAAAAA-----AGGAGGGYGLGSGGTSRGGLGGQAGAAAA--- 182
  |||||
Db 253 GQAGAAAAAGGAGGGYGLGGQAGGGYGLGSGG-AGRGGLGGQAGAAAAAGG 311
QY 183 -----AAAAAGGAGGGYGLGSGGTSRGGLGGQAGAAAA----- 220
  |||||
Db 312 AGGGLGGQAGGGAGAAAAAGGAGGGYGLGSGG-AGRGGLGGQAGAAAAAGGAG 370
QY 221 -----AAAAAGGAGGGYGLGSGGTSRGGLGGQAGAAAA--- 258
  |||||
Db 371 QGGYGLGSGGAGRGGGAGAAAAAGGAGGGYGLGSGG-AGRGGLGGQAGAAAA 429
QY 259 -----AAAAAGGAGGGYGLGSGG 279
  |||||
Db 430 GGAGGGYGLGSGGAGRGGGAGAAAAAGGAGGGYGLGSGG 474

RESULT 8
ID AAM50047
XX AAM50047 standard; Protein; 676 AA.
AC
XX AAM50047;
XX
DT 18-SEP-2002 (first entry)
DE N. clavipes spidroin synthetic homologue S01 protein #2.
KW Spidroin; spider; silk; fibre; film; membrane; wound; filter; S01.
OS Synthetic.
FH
FH Key Location/Qualifiers
FT Peptide 1..28
FT Protein /label= LeB4_signal_peptide
FT /note= "synthetic spidroin homologue S01"
FT Region 650..672
FT /note= "c-myc-tag"
FT Domain 673..676
FT /note= "ER retention signal"
XX
PN DE10113781-A1.
XX
XX 13-DEC-2001.
XX
XX 21-MAR-2001; 2001DE-1013781.
XX
PR 09-JUN-2000; 2000DE-1028212.
XX 24-OCT-2000; 2000DE-1053478.
XX
XX (IPKP-) IPK INST PFLANZENGENETIK & KULTURPFLANZE.
XX
```

```
PI Scheller J, Conrad U, Grosse F, Guehrs K;
XX WPI; 2002-123561/17.
XX
XX New DNA encoding synthetic spider silk protein, useful e.g. for closing
XX wounds, comprises modules that encode repeating units of spidroin
XX proteins
PS Example 1; Fig 10A; 88pp; German.
XX
XX This invention describes a novel DNA sequence, encoding a synthetic
XX spider silk protein, comprising modules, each comprising a group of
XX sequentially arranged oligonucleotides, each oligonucleotide encoding
XX a repeating unit of a spidroin protein. The synthetic protein has at
XX least 84% homology with the Nephila clavipes spidroin protein and is used
XX to produce synthetic fibres, films and/or membranes, particularly: (1)
XX for medical use, especially to close wounds and/or to support or cover
XX artificial organs; (ii) as adhesion surfaces for culturing cells; and
XX (iii) as filters. The synthetic proteins are very similar to native
XX spider silk proteins; can be prepared on a large scale and can be spun to
XX fibres with excellent mechanical properties (strength and elasticity).
XX Also they retain water solubility after long-term boiling in aqueous
XX solutions and since they are also soluble in organic solvents but
XX precipitated at high salt concentration, they are easily extracted and
XX purified. The modular construction of the invention facilitates
XX incorporation of additional peptide-encoding sequences, e.g. to simplify
XX purification or modulate solubility. This sequence represents a construct
XX composed of the LeB4 signal peptide, N. clavipes spidroin-1 synthetic
XX homologue S01, a c-myc-tag and an endoplasmic reticulum (ER)-retention
XX signal described in the invention.
SQ Sequence 676 AA;

Query Match 69.5%; Score 1031.5; DB 23; Length 676;
Best Local Similarity 59.8%; Pred. No. 1.4e-72;
Matches 242; Conservative 6; Mismatches 12; Indels 145; Gaps 13;

QY 11 GSMASGRGLGGQAGAAAA-----AAAAAGGAGGGY 44
  || :|||||
Db 107 GSGAGRGGLGGQAGAAAAAGGAGGGYGLGSGQAGRGGGQAGAAAAAGGAGGGY 166
QY 45 GGLGSGGTSRGGLGGQAGAAAAA-----AGGAGGGYGLGSGGTSRG 95
  ||||| :|||||
Db 167 GGLGSGG-AGRGGLGGQAGAAAAAGGAGGGYGLGGQAGGGYGLGSGG-AGRG 224
QY 96 LGGQAGAAAA-----AAAAAGGAGGGYGLGSGGTSRGGLG 135
  |||||
Db 225 LGGQAGAAAAAGGAGGGYGLGGQAGGGYGLGSGG-----AGRG 280
QY 136 GQAGAAAAA-----AGGAGGGYGLGSGGTSRGGLGGQAGAAAA--- 182
  |||||
Db 281 GQAGAAAAAGGAGGGYGLGGQAGGGYGLGSGG-AGRGGLGGQAGAAAAAGG 339
QY 183 -----AAAAAGGAGGGYGLGSGGTSRGGLGGQAGAAAA----- 220
  |||||
Db 340 AGGGLGGQAGGGAGAAAAAGGAGGGYGLGSGG-AGRGGLGGQAGAAAAAGGAG 398
QY 221 -----AAAAAGGAGGGYGLGSGGTSRGGLGGQAGAAAA--- 258
  |||||
Db 399 QGGYGLGSGGAGRGGGAGAAAAAGGAGGGYGLGSGG-AGRGGLGGQAGAAAA 457
QY 259 -----AAAAAGGAGGGYGLGSGG 279
  |||||
Db 458 GGAGGGYGLGSGGAGRGGGAGAAAAAGGAGGGYGLGSGG 502

RESULT 9
ID AAM50037
XX AAM50037 standard; Protein; 1255 AA.
XX
XX AAM50037;
XX
DT 18-SEP-2002 (first entry)
```

XX DE N. clavipes spideroin synthetic homologue SO1SO1 protein.  
XX KW Spideroin; spider; silk; fibre; film; membrane; wound; filter; SO1SO1.  
XX OS Synthetic.  
XX PN DE10113781-A1.  
XX PD 13-DEC-2001.  
XX PF 21-MAR-2001; 2001DE-1013781.  
XX PR 09-JUN-2000; 2000DE-1028212.  
XX PR 24-OCT-2000; 2000DE-1053478.  
XX PA (IPKP-) IPK INST PFLANZENGENETIK & KULTURPFLANZE.  
XX PI Scheller J, Conrad U, Grosse F, Guehrs K;  
XX XX WPI; 2002-123561/17.  
XX DR N-PSDB; ABL61038.  
XX PT New DNA encoding synthetic spider silk protein, useful e.g. for closing  
XX PT wounds, comprises modules that encode repeating units of spideroin  
XX PT proteins  
XX PS Claim 22; Page 29-34; 88pp; German.  
XX XX This invention describes a novel DNA sequence, encoding a synthetic  
XX CC spider silk protein, comprising modules, each comprising a group of  
XX CC sequentially arranged oligonucleotides, each oligonucleotide encoding  
XX CC a repeating unit of a spideroin protein. The synthetic protein has at  
XX CC least 84% homology with the Nephila clavipes spideroin protein and is used  
XX CC for medical use, especially to close wounds and/or to support or cover  
XX CC artificial organs; (ii) as adhesion surfaces for culturing cells; and  
XX CC (iii) as filters. The synthetic proteins are very similar to native  
XX CC spider silk proteins; can be prepared on a large scale and can be spun to  
XX CC fibres with excellent mechanical properties (strength and elasticity).  
XX CC Also they retain water solubility after long-term boiling in aqueous  
XX CC solutions and since they are also soluble in organic solvents but  
XX CC precipitated at high salt concentration, they are easily extracted and  
XX CC purified. The modular construction of the invention facilitates  
XX CC incorporation of additional peptide-encoding sequences, e.g. to simplify  
XX CC purification or modulate solubility. This sequence represents the  
XX CC synthetic N. clavipes spideroin-1 homologue SO1SO1 described in the  
XX CC invention.  
XX SQ Sequence 1255 AA;  
Query Match 69.5%; Score 1031.5; DB 23; Length 1255;  
Best Local Similarity 59.8%; Pred. No. 2.3e-72;  
Matches 242; Conservative 6; Mismatches 12; Indels 145; Gaps 13;  
QY 11 GSMASGRGGLGGGAGAAAA-----AAAAGAGAGGGY 44  
DB 704 GSGAGRGGLGGGAGAAAAAGAGAGGGYGGGGLGSGAGRGGGGAGAAAAAGAGGGY 763  
QY 45 GGLGSGQTSGRGLGGGAGAAAAA-----AGGAGGGYGGGSGQTSGRG 95  
DB 764 GGLGSGQ-AGRGGLGGGAGAAAAAGAGAGGGYGGGGLGGGAGGGYGGGGLGSGQ-AGRG 821  
QY 96 LGGGAGAAAA-----AAAAGAGAGGGYGGGGLGSGQTSGRGLG 135  
DB 822 LGGGAGAAAAAGAGAGGGGGLGGGAGAAAAAGAGAGGGYGGGGLGSGQ-----AGRG 877  
QY 136 GGGAGAAAAA-----AGGAGGGYGGGGLGSGQTSGRGLGGGAGAAAA-----182  
DB 878 GGGAGAAAAAGAGAGGGYGGGGLGGGAGGGYGGGGLGSGQ-AGRGGLGGGAGAAAAAGG 936  
QY 183 -----AAAAGAGAGGGYGGGGLGSGQTSGRGLGGGAGAAAA-----220  
|||||

DB 937 AGGGLGGGAGAGGAGAAAAAGAGGGYGGGGLGSGQ-AGRGGLGGGAGAAAAAGAG 995  
QY 221 -----AAAAGAGAGGGYGGGGLGSGQTSGRGLGGGAGAAAA-- 258  
|||||  
DB 996 QGGYGGGGLGSGAGRGGGGAGAAAAAGAGGGYGGGGLGSGQ-AGRGGLGGGAGAAAA 1054  
QY 259 -----AAAAGAGAGGGYGGGGLGSGQ 279  
|||||  
DB 1055 GGAGGGYGGGGLGSGAGRGGGGAGAAAAAGAGGGYGGGGLGSGQ 1099  
|||||

RESULT 10  
AAM50039  
ID AAM50039 standard; Protein; 1880 AA.  
XX AC AAM50039;  
XX DT 18-SEP-2002 (first entry)  
XX DE N. clavipes spideroin synthetic homologue SO1SO1SO1 protein.  
XX KW Spideroin; spider; silk; fibre; film; membrane; wound; filter; SO1SO1SO1.  
XX OS Synthetic.  
XX PN DE10113781-A1.  
XX PD 13-DEC-2001.  
XX PF 21-MAR-2001; 2001DE-1013781.  
XX PR 09-JUN-2000; 2000DE-1028212.  
XX PR 24-OCT-2000; 2000DE-1053478.  
XX PA (IPKP-) IPK INST PFLANZENGENETIK & KULTURPFLANZE.  
XX PI Scheller J, Conrad U, Grosse F, Guehrs K;  
XX WPI; 2002-123561/17.  
XX DR N-PSDB; ABL61040.  
XX PT New DNA encoding synthetic spider silk protein, useful e.g. for closing  
XX PT wounds, comprises modules that encode repeating units of spideroin  
XX PT proteins  
XX PS Claim 22; Page 38-45; 88pp; German.

This invention describes a novel DNA sequence, encoding a synthetic spider silk protein, comprising modules, each comprising a group of sequentially arranged oligonucleotides, each oligonucleotide encoding a repeating unit of a spideroin protein. The synthetic protein has at least 84% homology with the Nephila clavipes spideroin protein and is used to produce synthetic fibres, films and/or membranes, particularly: (i) for medical use, especially to close wounds and/or to support or cover artificial organs; (ii) as adhesion surfaces for culturing cells; and (iii) as filters. The synthetic proteins are very similar to native spider silk proteins; can be prepared on a large scale and can be spun to fibres with excellent mechanical properties (strength and elasticity). Also they retain water solubility after long-term boiling in aqueous solutions and since they are also soluble in organic solvents but precipitated at high salt concentration, they are easily extracted and purified. The modular construction of the invention facilitates incorporation of additional peptide-encoding sequences, e.g. to simplify purification or modulate solubility. This sequence represents the synthetic N. clavipes spideroin-1 homologue SO1SO1SO1 described in the invention.

XX SQ Sequence 1880 AA;

Query Match 69.5%; Score 1031.5; DB 23; Length 1880;  
Best Local Similarity 59.8%; Pred. No. 3.1e-72;  
Matches 242; Conservative 6; Mismatches 12; Indels 145; Gaps 13;









***This Page Blank (uspto)***



Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	1036	69.8	606	4	US-08-556-978B-21	Sequence 21, Appl
2	1036	69.8	606	4	US-09-247-806-4	Sequence 4, Appl
3	1015	68.4	651	4	US-08-556-978B-19	Sequence 19, Appl
4	1015	68.4	651	4	US-09-247-806-1	Sequence 1, Appl
5	1015	68.4	718	1	US-08-425-069-2	Sequence 2, Appl
6	1015	68.4	718	2	US-08-317-844B-2	Sequence 2, Appl
7	1015	68.4	747	3	US-09-034-177-3	Sequence 3, Appl
8	984	66.3	604	4	US-08-556-978B-63	Sequence 63, Appl
9	976	65.7	606	4	US-08-556-978B-23	Sequence 23, Appl
10	976	65.7	606	4	US-09-247-806-8	Sequence 8, Appl
11	970	65.3	606	4	US-09-247-806-6	Sequence 6, Appl
12	674	45.4	832	1	US-08-209-747-2	Sequence 2, Appl
13	674	45.4	832	1	US-08-458-298-2	Sequence 2, Appl
14	610.5	41.1	641	4	US-09-249-585A-3	Sequence 3, Appl
15	593.5	40.0	738	3	US-08-864-038A-3	Sequence 3, Appl
16	561	37.8	235	2	US-08-529-190B-1	Sequence 1, Appl
17	555.5	37.4	529	4	US-09-247-806-2	Sequence 2, Appl
18	554	37.3	595	1	US-08-425-069-4	Sequence 4, Appl
19	554	37.3	595	2	US-08-317-844B-4	Sequence 4, Appl
20	551	37.1	493	4	US-08-556-978B-59	Sequence 59, Appl
21	524	35.3	714	4	US-08-556-978B-61	Sequence 61, Appl
22	524	35.3	714	4	US-09-247-806-10	Sequence 10, Appl
23	520	35.0	1136	4	US-08-806-029-9	Sequence 9, Appl
24	520	35.0	1177	1	US-07-609-716-31	Sequence 31, Appl
25	520	35.0	1177	1	US-08-175-155-29	Sequence 29, Appl
26	520	35.0	1177	1	US-08-477-509B-64	Sequence 64, Appl
27	520	35.0	1177	2	US-08-707-237A-35	Sequence 35, Appl

```

Db      468  GQGAGAAAAAAGGAGGGYGGGLGSQSGAGGGYGGLGSQSGAGRGCGQAGAAAAAAAGGAGQG 271
Qy      233  GYGGLGSQGTSGRGLGGCGAGAAA-----AAAAAAAAAGGAGQG 271
          ||||| :|||||:|||||:|||||
Db      528  GYGGLGSQG-AGRGLGGCGAGAAAAAAGGAGGGGLGSQSGAGAAAAAAGGAGQG 586
          ||||| :|||||:|||||:|||||
Qy      272  YGGLGSQG 279
          |||||
Db      587  YGGLGSQG 594

RESULT 3
US-08-556-978B-19
; Sequence 19, Application US/08556978B
; Patent No. 6268169
; GENERAL INFORMATION:
; APPLICANT: FARNESTOCK, STEPHEN F.
; TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED
; TYPE OF INVENTION: SPIDER SILK ANALOGS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD FOR WINDOWS 95
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/556,978B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/077,600
; FILING DATE: JUNE 15, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA AXAMETHY
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CR-9389-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 651 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-556-978B-19

Query Match           68.4%; Score 1015; DB 4; Length 651;
Best Local Similarity 59.2%; Pred. No. 3e-75;
Matches 238; Conservative 8; Mismatches 14; Indels 142; Gaps

Qy      11  GSMASGRGGLGGCGACAAAA-----AAAAAAAAAGGAGGGYGC 45
          || :|||:|||||:|||||
Db      93  GSQCAGRGLGGCGAGAAAAAAGGAGGGYGGGLGNQAGRCQQGAAAAAACGAGCGGYG 152
          |||||:|||||:|||||:|||||

Qy      46  GLGSQGTSGRGLGGCGAGAAAAA-----AGGAGGGYGGGLGSQGTSGRGL 96
          |||||:|||||:|||||:|||||
Db      153  GLGSQG-AGRGLGGCGAGAAAAAAGGAGGGYGGGLGGCGAGCGYGGGLSQG-ACRGL 210
          |||||:|||||:|||||:|||||

Qy      97  GGCGAGAAAA-----AAAGGACGGYGGGLGSQGTSGRGLG 135
          |||||:|||||:|||||:|||||
Db      211  GGCGAGAAAAAAGGAGGGYGGGLGGCGAGAGASAAAAAGGAGGGYGGGLSQG----AGRG 266
          |||||:|||||:|||||:|||||
Qy      136  GCGAGAAAAA-----AGGAGGGYGGGLGSQGTSGRGLGGCGAGAAAA--- 182
          |||||:|||||:|||||:|||||

```















Result No.	Score	Query			DB	ID	Description
		Match	Length				
1	1036	69.8	606	10	US-09-861-597-4	Sequence 4, Appli	
2	1015	68.4	651	10	US-09-861-597-1	Sequence 1, Appli	
3	976	65.7	606	10	US-09-861-597-8	Sequence 8, Appli	
4	970	65.3	606	10	US-09-861-597-6	Sequence 6, Appli	
5	555.5	37.4	529	10	US-09-861-597-2	Sequence 2, Appli	
6	524	35.3	714	10	US-09-861-597-7	Sequence 10, Appl	
7	466.5	31.4	357	10	US-09-864-761-35807	Sequence 35807, A	
8	433	29.2	201	10	US-09-848-990-22	Sequence 22, Appl	
9	433	29.2	201	10	US-09-760-364-14	Sequence 14, Appl	
10	431.5	28.1	200	10	US-09-798-584-18	Sequence 18, Appl	
11	431.5	29.1	200	10	US-09-367-624-19	Sequence 19, Appl	
12	431.5	29.1	200	10	US-09-998-667-18	Sequence 18, Appl	
13	421	28.4	101	10	US-09-861-597-5	Sequence 7, Appli	
14	412	27.7	101	10	US-09-861-597-7	Sequence 5, Appli	
15	407.5	27.4	1894	12	US-10-052-586-97	Sequence 97, Appl	
16	404.5	27.2	101	10	US-09-861-597-3	Sequence 3, Appli	
17	402	27.1	283	10	US-09-864-761-36720	Sequence 36720, A	
18	370	24.9	1400	10	US-09-879-957-37	Sequence 37, Appli	
19	343.5	23.1	1497	10	US-09-060-8548-2	Sequence 2, Appli	



```

Query Match      37.4%; Score 555.5; DB 10; Length 529;
Best Local Similarity 44.1%; Pred. No. 8e-28;
Matches 167; Conservative 11; Mismatches 90; Indels 111; Gaps 18;

QY 11 GSMASGRGGLGGCGAG-----AAANAAAAAG---CAGC---GGYG----- 45
DB 2 GGYGPGQQPGGYGPGQQGSPGSAANAAAAAANGPGGYGPGQQGPGYGGQQGPGRY 61
QY 46 GLGSQGTSGRG-----GLGGQGAG-----AAAAAANA 74
DB 62 GPGQGPSPGSAANAAAAAGSQOQPGGYGPRQOQPGGYGQOQPGSPGSAANASA 121
QY 75 GGACQGGYG-----GLGSQGTSGRG-----GLGGQGAGANAAAAAGGAGQGGYG 121
DB 122 ESGGPGGYGPGQPGGYGPGQQGPGGYGPGQQGPGSPGSAANAAAAAGSPGQOQPGYG 181
QY 122 GLGSQGTSGRG---GLGGQGAGANAAAAAGAGGQGYGGLGSQGTSGRG---GLG 173
DB 182 P-GQQGPGGYGPGQQGSPGSAANAAAAAGSPGQOQPGGYG-PGQGPGGYGPGQOGLS 239
QY 174 QGAGAAAA-----AAAAAGAGAGGQGYGGLGSQGTSGRG---GLGSQGTSGRG 208
DB 240 GPGSAANAAAAAGPGQOQPGGYGPGQOQPGSPGSAANAAAAAGPGGYGPGQOQPGYGP 299
QY 209 GLGQOQAGANAAAAAGAGGQGYGGLGSQGTSGRG---GLGGQGAGANAAAAANA 264
DB 300 QQGPSPGAGSAANAAAAAGPGQOGLGYP-GQQGPGGYGPGQOQPGGYGPGSASAAAAAG 358
QY 265 -GGAGQGYGGLGSQGTSG 282
DB 359 PGQOQPGGYG-PGQOQPSG 376

RESULT 6
US-09-861-597-10
; Sequence 10, Application US/09861597
; Patent No. US20020064539A1
; GENERAL INFORMATION:
; APPLICANT: PHILLIPPE, Michel
; APPLICANT: GARSON, Jean-Claude
; APPLICANT: ARRAUDEAU, Jean-Pierre
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
; TITLE OF INVENTION: ANALOG
; FILE REFERENCE: 6388-0365-0
; CURRENT APPLICATION NUMBER: US/09/861.597
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: 09/247,806
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: FR 98/01614
; PRIOR FILING DATE: 1998-02-11
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 714
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:protein
US-09-861-597-10

Query Match      35.3%; Score 524; DB 10; Length 714;
Best Local Similarity 49.2%; Pred. No. 8e-26;
Matches 155; Conservative 12; Mismatches 107; Indels 41; Gaps 17;

QY 10 HGSMASGRGGLG-GQ---GAGAAAAAANAAG---CAGOGGYGGLGSQGTSGRGGLG- 59
DB 296 YGPGQOQPGGYGPGQOQPGSPGSAANAAAAAGPGGYGPGQOQPGGYG-PGQOQPGGYGP 354
QY 60 QGAGAAAAAANAAGAGGQGYGGLGSQGTSGRGGLG-GQAGAAAAAANAAGAGAGQ 118
DB 355 GQOQPGSPGSAANAAAAAGPGQOQPGGYG-PGQOQPGGYGPGQOQPGSPGSAANAAAAAGPG 413

```



[illegible]

US-09-967-624-19  
 ; Sequence No. 19, Application US/03967624  
 ; Patent No. US2002014325A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Liao, X. Charlene  
 ; APPLICANT: Chu, Peter  
 ; APPLICANT: Pardo, Jorge



```
Db          95 YGGLGSQ 101
|||||
RESULT 14
US-09-861-597-5
; Sequence 5, Application US/09861597
; Patent No. US20020064539A1
; GENERAL INFORMATION:
; APPLICANT: PHILIPPE, Michel
; APPLICANT: ARRAUDEAU, Jean-Pierre
; APPLICANT: ARRAUDEAU, Jean-Pierre
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
; TITLE OF INVENTION: ANALOG
; FILE REFERENCE: 6388-0365-0
; CURRENT APPLICATION NUMBER: US/09/861.597
; CURRENT FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: 09/247,806
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: FR 98/01614
; PRIOR FILING DATE: 1998-02-11
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Nephila clavipes
US-09-861-597-5

Query Match          27.7%   Score 412; DB 10; Length 101;
Best Local Similarity 73.6%; Pred. No. 1.2e-19;
Matches 92; Conservative 1; Mismatches 6; Indels 26; Gaps 5;

QY 152 GAGGGYGGGSGTSGRGLGGGAGAGAAAAAGAGGGYGGGSGTSGRGLG 211
|||||
Db 3 GAGGGYGGGSGG-AGRGLGGGAG-AAAAAGAGGAG-AGGAGGAG-AGGAGGAGG 44
|||||

QY 212 GQAGAGAAAAAGAGAGGGYGGGSGTSGRGLGGGAGAGAAAAAGAGAGG 271
|||||
Db 45 SQAGAGAGAAAAAGAGGGYGGGSGG-AGRGGGAG-AGRGGGAG-AGRGGGAGG 96
|||||

QY 272 YGGLG 276
|||||
Db 97 YGGLG 101
|||||

RESULT 15
US-10-052-586-97
; Sequence 97, Application US/10052586
; Patent No. US20020127584A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C1
; CURRENT APPLICATION NUMBER: US/10/052,586
; CURRENT FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
```

;

PRIOR FILING DATE: 1997-10-17

PRIOR APPLICATION NUMBER: 60/063120

;

PRIOR FILING DATE: 1997-10-24

PRIOR APPLICATION NUMBER: 60/063121

;

PRIOR FILING DATE: 1997-10-24

PRIOR APPLICATION NUMBER: 60/063486

;

PRIOR FILING DATE: 1997-10-21

PRIOR APPLICATION NUMBER: 60/063540

;

PRIOR FILING DATE: 1997-10-28

PRIOR APPLICATION NUMBER: 60/063541

;

PRIOR FILING DATE: 1997-10-28

PRIOR APPLICATION NUMBER: 60/063544

;

PRIOR FILING DATE: 1997-10-28

PRIOR APPLICATION NUMBER: 60/063564

;

PRIOR FILING DATE: 1997-10-28

PRIOR APPLICATION NUMBER: 60/063734

;

PRIOR FILING DATE: 1997-10-29

PRIOR APPLICATION NUMBER: 60/063870

;

PRIOR FILING DATE: 1997-10-31

PRIOR APPLICATION NUMBER: 60/064103

;

PRIOR FILING DATE: 1997-10-31

PRIOR APPLICATION NUMBER: 60/065311

;

PRIOR FILING DATE: 1997-11-13

PRIOR APPLICATION NUMBER: 60/066120

;

PRIOR FILING DATE: 1997-11-21

PRIOR APPLICATION NUMBER: 60/066466

;

PRIOR FILING DATE: 1997-11-24

PRIOR APPLICATION NUMBER: 60/066772

;

PRIOR FILING DATE: 1997-11-24

PRIOR APPLICATION NUMBER: 60/069335

;

PRIOR FILING DATE: 1997-12-11

PRIOR APPLICATION NUMBER: 60/069425

;

PRIOR FILING DATE: 1997-12-12

PRIOR APPLICATION NUMBER: 60/069870

;

PRIOR FILING DATE: 1997-12-17

PRIOR APPLICATION NUMBER: 60/068017

;

PRIOR FILING DATE: 1997-12-18

PRIOR APPLICATION NUMBER: 60/077450

;

PRIOR FILING DATE: 1998-03-10

PRIOR APPLICATION NUMBER: 60/077632

;

PRIOR FILING DATE: 1998-03-11

PRIOR APPLICATION NUMBER: 60/077649

;

PRIOR FILING DATE: 1998-03-11

PRIOR APPLICATION NUMBER: 60/078886

;

PRIOR FILING DATE: 1998-03-20

PRIOR APPLICATION NUMBER: 60/078939

;

PRIOR FILING DATE: 1998-03-20

PRIOR APPLICATION NUMBER: 60/079664

;

PRIOR FILING DATE: 1998-03-27

PRIOR APPLICATION NUMBER: 60/079786

;

PRIOR FILING DATE: 1998-03-27

PRIOR APPLICATION NUMBER: 60/080107

;

PRIOR FILING DATE: 1998-03-31

PRIOR APPLICATION NUMBER: 60/080194

;

PRIOR FILING DATE: 1998-03-31

PRIOR APPLICATION NUMBER: 60/080327

;

PRIOR FILING DATE: 1998-04-01

PRIOR APPLICATION NUMBER: 60/080333

;

PRIOR FILING DATE: 1998-04-01

PRIOR APPLICATION NUMBER: 60/081049

;

PRIOR FILING DATE: 1998-04-08

PRIOR APPLICATION NUMBER: 60/081070

;

PRIOR FILING DATE: 1998-04-08

PRIOR APPLICATION NUMBER: 60/081195

;

PRIOR FILING DATE: 1998-04-09

PRIOR APPLICATION NUMBER: 60/081838

;

PRIOR FILING DATE: 1998-04-15

PRIOR APPLICATION NUMBER: 60/082568

;

PRIOR FILING DATE: 1998-04-21

PRIOR APPLICATION NUMBER: 60/082569

;

PRIOR FILING DATE: 1998-04-21

PRIOR APPLICATION NUMBER: 60/082704

;

PRIOR FILING DATE: 1998-04-22

;	PRIOR APPLICATION NUMBER: 60/082797	;
;	PRIOR FILING DATE: 1998-04-22	;
;	PRIOR APPLICATION NUMBER: 60/083322	;
;	PRIOR FILING DATE: 1998-04-28	;
;	PRIOR APPLICATION NUMBER: 60/083495	;
;	PRIOR FILING DATE: 1998-04-29	;
;	PRIOR APPLICATION NUMBER: 60/083496	;
;	PRIOR FILING DATE: 1998-04-29	;
;	PRIOR APPLICATION NUMBER: 60/083499	;
;	PRIOR FILING DATE: 1998-04-29	;
;	PRIOR APPLICATION NUMBER: 60/083559	;
;	PRIOR FILING DATE: 1998-04-29	;
;	PRIOR APPLICATION NUMBER: 60/084366	;
;	PRIOR FILING DATE: 1998-05-05	;
;	PRIOR APPLICATION NUMBER: 60/084414	;
;	PRIOR FILING DATE: 1998-05-06	;
;	PRIOR APPLICATION NUMBER: 60/084639	;
;	PRIOR FILING DATE: 1998-05-07	;
;	PRIOR APPLICATION NUMBER: 60/084640	;
;	PRIOR FILING DATE: 1998-05-07	;
;	PRIOR APPLICATION NUMBER: 60/084643	;
;	PRIOR FILING DATE: 1998-05-07	;
;	PRIOR APPLICATION NUMBER: 60/085573	;
;	PRIOR FILING DATE: 1998-05-15	;
;	PRIOR APPLICATION NUMBER: 60/085579	;
;	PRIOR FILING DATE: 1998-05-15	;
;	PRIOR APPLICATION NUMBER: 60/085580	;
;	PRIOR FILING DATE: 1998-05-15	;
;	PRIOR APPLICATION NUMBER: 60/085582	;
;	PRIOR FILING DATE: 1998-05-15	;
;	PRIOR APPLICATION NUMBER: 60/085700	;
;	PRIOR FILING DATE: 1998-05-15	;
;	PRIOR APPLICATION NUMBER: 60/086023	;
;	PRIOR FILING DATE: 1998-05-18	;
;	PRIOR APPLICATION NUMBER: 60/086392	;
;	PRIOR FILING DATE: 1998-05-22	;
;	PRIOR APPLICATION NUMBER: 60/086486	;
;	PRIOR FILING DATE: 1998-05-22	;
;	PRIOR APPLICATION NUMBER: 60/087098	;
;	PRIOR FILING DATE: 1998-05-28	;
;	PRIOR APPLICATION NUMBER: 60/087208	;
;	PRIOR FILING DATE: 1998-05-28	;
;	PRIOR APPLICATION NUMBER: 60/087609	;
;	PRIOR FILING DATE: 1998-06-02	;
;	PRIOR APPLICATION NUMBER: 60/087759	;
;	PRIOR FILING DATE: 1998-06-02	;
;	PRIOR APPLICATION NUMBER: 60/087827	;
;	PRIOR FILING DATE: 1998-06-03	;
;	PRIOR APPLICATION NUMBER: 60/088025	;
;	PRIOR FILING DATE: 1998-06-04	;
;	PRIOR APPLICATION NUMBER: 60/088028	;
;	PRIOR FILING DATE: 1998-06-04	;
;	PRIOR APPLICATION NUMBER: 60/088029	;
;	PRIOR FILING DATE: 1998-06-04	;
;	PRIOR APPLICATION NUMBER: 60/088033	;
;	PRIOR FILING DATE: 1998-06-04	;
;	PRIOR APPLICATION NUMBER: 60/088167	;
;	PRIOR FILING DATE: 1998-06-05	;
;	PRIOR APPLICATION NUMBER: 60/088202	;
;	PRIOR FILING DATE: 1998-06-05	;
;	PRIOR APPLICATION NUMBER: 60/088212	;
;	PRIOR FILING DATE: 1998-06-05	;
;	PRIOR APPLICATION NUMBER: 60/088217	;
;	PRIOR FILING DATE: 1998-06-05	;
;	PRIOR APPLICATION NUMBER: 60/088326	;
;	PRIOR FILING DATE: 1998-06-04	;
;	PRIOR APPLICATION NUMBER: 60/088655	;
;	PRIOR FILING DATE: 1998-06-09	;
;	PRIOR APPLICATION NUMBER: 60/088722	;
;	PRIOR FILING DATE: 1998-06-10	;
;	PRIOR APPLICATION NUMBER: 60/088738	;
;	PRIOR FILING DATE: 1998-06-10	;
;	PRIOR APPLICATION NUMBER: 60/088740	;

```

, PRIOR FILING DATE: 1998-06-10
, PRIOR APPLICATION NUMBER: 60/008811
, PRIOR FILING DATE: 1998-06-10
, PRIOR APPLICATION NUMBER: 60/008824
, PRIOR FILING DATE: 1998-06-10
, PRIOR APPLICATION NUMBER: 60/008825
, PRIOR FILING DATE: 1998-06-10
, PRIOR APPLICATION NUMBER: 60/008826
, PRIOR FILING DATE: 1998-06-10
, PRIOR APPLICATION NUMBER: 60/008861
, PRIOR FILING DATE: 1998-06-11
, PRIOR APPLICATION NUMBER: 60/008863
, PRIOR FILING DATE: 1998-06-11
, PRIOR APPLICATION NUMBER: 60/008876
, PRIOR FILING DATE: 1998-06-11
, PRIOR APPLICATION NUMBER: 60/008900
, PRIOR FILING DATE: 1998-06-12
, PRIOR APPLICATION NUMBER: 60/008910
, PRIOR FILING DATE: 1998-06-12
, PRIOR APPLICATION NUMBER: 60/008912
, PRIOR FILING DATE: 1998-06-16
, PRIOR APPLICATION NUMBER: 60/008951
, PRIOR FILING DATE: 1998-06-16
, PRIOR APPLICATION NUMBER: 60/008953
, PRIOR FILING DATE: 1998-06-17
, PRIOR APPLICATION NUMBER: 60/008958
, PRIOR FILING DATE: 1998-06-17
, PRIOR APPLICATION NUMBER: 60/008953
, PRIOR FILING DATE: 1998-06-17
, PRIOR APPLICATION NUMBER: 60/008908

```

Query Match 27.4%; Score 407.5; DB 12; Length 1894;  
Best Local Similarity 35.0%; Pred. No. 1.8e-18;  
Matches 133; Conservative 12; Mismatches 122; Indels 113; Gaps 13;

Search completed: December 18, 2002, 16:38:50  
Job time : 8.01995 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:13:49 ; Search time 8.59824 Seconds  
(without alignments)  
3253.588 Million cell updates/sec

Title: US-09-490-291-2

Perfect score: 1485

Sequence: 1 MRGSHHHHHGSMASGRGL.....YGLGSGQGTSGIRPAKLN 291

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_73:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1015	68.4	718	2 A36068	major ampullate fi
2	617.5	41.6	2639	2 T31328	fibroin - Chinese
3	610.5	41.1	641	1 Q0BE31	nuclear antigen EB
4	556.5	37.5	1901	2 F70806	hypothetical glyci
5	556	37.4	1489	2 D70807	hypothetical glyci
6	554	37.3	627	2 A44112	spidroin 2, dragli
7	544.5	36.7	1381	2 E70806	hypothetical glyci
8	543.5	36.6	1079	2 E70807	hypothetical glyci
9	538	36.2	1538	2 H70846	hypothetical glyci
10	535	36.0	749	2 A70812	hypothetical glyci
11	525.5	35.4	1660	2 A70869	hypothetical glyci
12	525	35.4	384	1 A26099	glycine-rich cell
13	519.5	35.0	853	2 A70896	hypothetical glyci
14	516.5	34.8	778	2 F70963	hypothetical glyci
15	514	34.6	1306	2 A70934	hypothetical glyci
16	512.5	34.5	767	2 E70895	hypothetical glyci
17	512	34.5	484	2 G70846	hypothetical glyci
18	507.5	34.2	923	2 E70820	hypothetical glyci
19	503	33.9	491	2 D70916	hypothetical glyci
20	503	33.9	731	2 C70974	hypothetical glyci
21	499	33.6	714	2 A70807	hypothetical glyci
22	498.5	33.6	882	2 B70812	hypothetical glyci
23	496.5	33.4	465	1 S01820	glycine-rich cell
24	496	33.4	603	2 A70770	hypothetical glyci
25	495	33.3	576	2 A70900	hypothetical glyci
26	495	33.3	1329	2 E70917	hypothetical glyci
27	494	33.3	667	2 A70893	hypothetical glyci
28	494	33.3	914	2 H70987	hypothetical glyci
29	490	33.0	338	1 KNMU	glycine-rich cell

30	490	33.0	801	2 F70824	hypothetical glyci
31	489.5	33.0	291	1 S31415	glycine-rich prote
32	489.5	33.0	396	2 T49109	glycine-rich prote
33	489.5	33.0	618	2 A70989	hypothetical glyci
34	489	32.9	498	2 C70720	hypothetical glyci
35	487.5	32.8	694	2 F70868	hypothetical glyci
36	487	32.8	496	2 H70839	hypothetical glyci
37	486.5	32.8	741	2 G70917	hypothetical glyci
38	486	32.7	1011	2 F70620	hypothetical glyci
39	484	32.6	615	2 H70589	hypothetical glyci
40	481	32.4	591	2 B70523	hypothetical glyci
41	480.5	32.4	532	2 F70580	hypothetical glyci
42	480.5	32.4	783	2 E70824	hypothetical glyci
43	479	32.3	594	2 G70545	hypothetical glyci
44	477.5	32.2	606	2 H70816	hypothetical glyci
45	477	32.1	837	2 E70835	hypothetical glyci

#### ALIGNMENTS

##### RESULT 1

A36068

major ampullate fibroin protein - orb spider (Nephila clavipes) (fragment)

C:Species: Nephila clavipes

C:Date: 08-Mar-1991 #sequence\_revision 13-Jan-1993 #text\_change 09-Sep-1997

C:Accession: A36068

R:Xu, M.; Lewis, R.V.

Proc. Natl. Acad. Sci. U.S.A. 87, 7120-7124, 1990

A:Title: Structure of a protein superfiber: spider dragline silk.

A:Reference number: A36068; MUID:90384959; PMID:2402494

A:Accession: A36068

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-718 <XUA>

A:Cross-references: GB:M37137; NID:g159711; PID:g159712

A:Note: the authors translated the codon GGT for residue 292 as Gln, GTA for residue

Query Match 68.4%; Score 1015; DB 2; Length 718;

Best Local Similarity 59.2%; Pred. No. 1.9e-53;

Matches 238; Conservative 8; Mismatches 14; Indels 142; Gaps 13;

Qy	11	GSMASGRGLGGQGAGAAAA	-----	-----	AAAAAGGAGGGG	45
Db	93	GSQAGRGRLGGQGAGAAAA	AAAAAGGAGGGG	GGGGLGNQAGRG	GGGQAGGAGGGG	152
Qy	46	GLGSGQGTSGRGLGGQGAG	AAAAA	-----	AGGAGGGG	96
Db	153	GLGSGQ-AGRGGLGGQGAG	AAAAAGGAGGGG	GGGGLGGQGAG	GGGGLGGSGQ-AGRGGL	210
Qy	97	GGQGAGAAAAA	-----	-----	AAAGCAGGGG	135
Db	211	GGQGAGAAAAA	AGGAGGGG	GGGGLGGQGAG	AGGAGGGG	266
Qy	136	GGAGAAAAA	-----	-----	AGGAGGGG	182
Db	267	GEGAGAAAAA	AGGAGGGG	GGGGLGGSGQ	AGRGGLGGGAGAAAGG	325
Qy	183	-----	-----	-----	AAAAAGGAGGGG	220
Db	326	QGGLGGQGAGAG	AAAAAGGAGGGG	GGGGLGGSGQ	AGRGGLGGGAGAAAGG	384
Qy	221	-----	-----	-----	AAAAAGGAGGGG	259
Db	385	GGYGLGSGQAGRG	QAGAAAAA	AGGAGGGG	GGGGLGGSGQ	443
Qy	260	-----	-----	-----	AAAAAGGAGGGG	279
Db	444	GGAGGGG	GGGLGNQAGRG	GGGAGAAAGG	AGGAGGGG	485

##### RESULT 2

T31328

R;Sample, J.; Hummel, M.; Braun, D.; Birkenbach, M.; Kieff, E.  
Proc. Natl. Acad. Sci. U.S.A. 83, 5096-5100, 1986  
A>Title: Nucleotide sequences of mRNAs encoding Epstein-Barr virus nuclear proteins:  
A:Reference number: S42440; MUID:86259739; PMID:3460083  
A:Accession: S42440  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-66 <SAM>  
A:Cross-references: EMBL:ML13941; NID:g330399; PIDN:AAA45889.1; PID:g555157  
C:Superfamily: Epstein-Barr virus nuclear antigen  
C:Keywords: DNA binding; transcription regulation

Query Match            41.1%; Score 610.5; DB 1; Length 641;  
Best Local Similarity 48.6%; Pred. No. 1.2e-29;  
Matches 143; Conservative 14; Mismatches 104; Indels 33; Gaps 10;

QY    2 RGSHHHHSMSGRGLG-GQGAGAAAAAAAGAGAG-QGYGGLSGQTSGRGLG 59  
      :| | | | | : | | | | | : | | | | | : | | | | | : | |  
Db    83 KGTH-----GCTGAGAGGAGAGGAGGAGGGAGGGAGGAGGAGGAGA 130  
  
QY    60 GQAGAAAAAAAGAGAG-QGYGGLSGQTSRGLGCGACAAAAAAAG---GA 115  
      : | | | | | : | | | | | : | | | | | : | | | | | : | |  
Db    131 GAGGAGGAGAGGAGGAGGAGGAGGAGG--GAGGAGGAGGAGGAGGAGGAGG 188  
  
QY    116 QGGYGGLSGQTSRGLGAG-QAGAAAAAAAGAGAGAGGAGGAGGAGGAGG 174  
      || | | | | : | | | | | : | | | | | : | | | | | : | |  
Db    189 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 244  
  
QY    175 QGAGAAAAAAAGAGAGGQGYGGLSGQTSG-----RGGLGGAGAAAAAAAGGAG 230  
      || | | | | : | | | | | : | | | | | : | | | | | : ||||  
Db    245 AGAGGAGGAGGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 304  
  
QY    231 QGGYGGLSGQTSRGLGCGACAAAAAAAGAGAGGAGGQGYGGLSGQTSGIR 284  
      || | | | | : | | | | | : | | | | | : | | | | | : | |  
Db    305 AGGAGGAGGAGGAGGAGGAGGAGG----RGRGGGRRGGSGRGRGGRR 354

RESULT 4  
F70806  
hypothetical glycine-rich protein RV3508 - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: F70806  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, P.  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence  
A:Reference number: A70500; MUID:9825987; PMID:9634230  
A:Accession: F70806  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1901 <COL>  
A:Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAA17745.1; PID:g29292  
A:Experimental source: strain H37RV  
C:Genetics:  
C:Gene: RV3508  
C:Superfamily: collagen alpha 1(IV) chain

Query Match            37.5%; Score 556.5; DB 2; Length 1901;  
Best Local Similarity 44.9%; Pred. No. 4e-26;  
Matches 140; Conservative 13; Mismatches 112; Indels 47; Gaps 14;

QY    14 ASRGGGLGGGAGAAAAAAAGAGGAGGQGYGGLSGQTSGR-----GGILGGOGAG 64  
      | | | | | | : | | | | | : | | | | | : | | | | | : | |  
Db    416 AGGAGGVGGGNAAGINGSGGAGGTGGGGAGGAGGAGGAGGAGGAGGAGTGGN 475  
  
QY    65 AAAAAAAAGGAGCGGY-----GGLSGQTSRGLGCGACAAAAAAAGAGAG--GG 119  
      | | | | | | : | | | | | : | | | | | : | | | | | : | |  
Db    476 AGAGGAGCAATGCTGTGTVGAAGKAGTCGTGGGAGGAGSAGTATATGATGCTGTSFG 535  
  
QY    120 YGGL-----GSGQTSRGLGCGACAAAAAAAGAGAGO-----GGYGGLSGQTSGR 169









Db 1201 GFGDGLDASSG 1212

Search completed: December 18, 2002, 16:21:57  
Job time : 12.5982 secs

***This Page Blank (uspto)***



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:06:04 ; Search time 4.46662 Seconds  
(without alignments)  
2702.183 Million cell updates/sec

Title: US-09-490-291-2

Perfect score: 1485

Sequence: 1 MRCGSHHHHGHGSMASGRGGL.....YGLGSGQTGIRPAKLN 291

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1015	68.4	747	1 SPDL_NEPCL	P19837 nephila cla
2	610.5	41.1	641	1 EBN1_EBV	P03211 epstein-bar
3	556.5	37.5	1901	1 Y208_MYCTU	O53553 mycobacteri
4	554	37.3	627	1 SPDL_NEPCL	P46804 nephila cla
5	525	35.4	384	1 GRP1_PETHY	P09789 petunia hyb
6	516.5	34.8	778	1 Y034_MYCTU	P71933 mycobacteri
7	508.5	34.2	5263	1 FBOH_BOMMO	P05790 bombyx mori
8	496.5	33.4	465	1 GRP2_PHAVU	P10496 phaseolus v
9	496	33.4	603	1 YD25_MYCTU	O10637 mycobacteri
10	494	33.3	914	1 WA22_MYCTU	O06794 mycobacteri
11	490	33.0	338	1 GRP_ARATH	P27483 arabidopsis
12	490	33.0	801	1 Y747_MYCTU	O53810 mycobacteri
13	489	32.9	498	1 Y118_MYCTU	O50615 mycobacteri
14	472	31.8	860	1 ELS_MOUSE	P54320 mus musculu
15	467	31.4	957	1 Y278_MYCTU	P56877 mycobacteri
16	463	31.2	543	1 YP91_MYCTU	Q50630 mycobacteri
17	453.5	30.5	864	1 ELS_RAT	P18165 mus musculu
18	444.5	29.9	463	1 Y468_MYCTU	P04985 bos taurus
19	443.5	29.9	491	1 YK98_MYCTU	P07916 gallus gall
20	438	29.5	252	1 GRP1_PHAVU	P23490 homo sapien
21	418.5	28.2	515	1 Y140_MYCTU	P10495 phaseolus v
22	409.5	27.6	481	1 LORI_MOUSE	O50594 mycobacteri
23	394.5	26.6	747	1 ELS_BOVIN	P18165 mus musculu
24	385.5	26.0	750	1 ELS_CHICK	P04985 bos taurus
25	378	25.5	316	1 LORI_HUMAN	P07916 gallus gall
26	374	25.2	730	1 ELS_HUMAN	P10495 phaseolus v
27	363	24.4	375	1 SANT_PLAFV	P10495 phaseolus v
28	357.5	24.1	2038	1 FSH_DROME	P13709 drosophila
29	341.5	23.0	401	1 CSP_PLACC	P08674 plasmodium
30	332.5	22.4	1733	1 VNVA_PRVKA	P33485 pseudorabie
31	330.5	22.3	1380	1 DDG9_MOUSE	O70133 mus musculu
32	328	22.1	398	1 CSP_PLACC	P08673 plasmodium
33	325	21.9	1356	1 CA21_ONCMY	O93484 oncorhynch

#### ALIGNMENTS

RESULT 1

ID	SPDL_NEPCL	STANDARD	PRT	747 AA
DT	01-FEB-1991 (Rel. 17, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Spidroin 1 (Dragline silk fibroin 1) (Fragment).			
OS	Nephila clavipes (Orb spider).			
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;			
OC	Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.			
OX	NCBI_TaxID=6915;			
RN	[1]			
RP	SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.			
RX	MEDLINE=90384959; PubMed=240494;			
RA	Xu M., Lewis R.V.;			
RT	"Structure of a protein superfiber: spider dragline silk.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).			
RN	[2]			
RP	SEQUENCE OF 653-747 FROM N.A.			
RX	MEDLINE=94165058; PubMed=8120021;			
RA	Beckwith R., Arcidiacono S.;			
RT	"Sequence conservation in the C-terminal region of spider silk proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and Araneus bicentenarius (Araneidae).";			
RL	J. Biol. Chem. 269:6661-6663(1994).			
CC	-1- FUNCTION: Spiders major ampullate silk possesses unique characteristics of strength and elasticity. Fibroin consists of pseudocrystalline regions of antiparallel beta-sheet interspersed with elastic amorphous segments.			
CC	-1- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 2, OF THE DRAGLINE SILK.			
CC	-1- SUBCELLULAR LOCATION: Extracellular.			
CC	-1- DOMAIN: Highly repetitive protein characterized by regions of polyalanine and glycine-rich repeating units.			
CC	-1- SIMILARITY: BELONGS TO THE SILK FIBROIN FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; M37137; AAA29380.1; -			
DR	EMBL; U03848; AAB60212.1; -			
DR	PIR; A36068; A36068.			
DR	Silk; Repeat.			
FT	NON_TER	1	655	25 X APPROXIMATE TANDEM REPEATS.
FT	DOMAIN	1	25	1.
FT	REPEAT	26	38	2.
FT	REPEAT	39	66	3.
FT	REPEAT	67	96	4.
FT	REPEAT	97	130	5.
FT	REPEAT	131	158	6.

```
FT REPEAT 159 191 7.
FT REPEAT 192 204 8.
FT REPEAT 205 235 9.
FT REPEAT 236 262 10.
FT REPEAT 263 292 11.
FT REPEAT 293 305 12.
FT REPEAT 306 333 13.
FT REPEAT 334 360 14.
FT REPEAT 361 394 15.
FT REPEAT 395 424 16.
FT REPEAT 425 458 17.
FT REPEAT 459 485 18.
FT REPEAT 486 512 19.
FT REPEAT 513 525 20.
FT REPEAT 526 555 21.
FT REPEAT 556 582 22.
FT REPEAT 583 612 23.
FT REPEAT 613 642 24.
FT REPEAT 643 655 25.
FT CONFLICT 662 662 V -> L (IN REF. 1).
FT CONFLICT 672 672 S -> T (IN REF. 1).
FT CONFLICT 695 747 NPGSGCDVLIOALLLEWVSALIQILGSSSIGOVNYSAGQA
      TOIVGOSVYQAL -> ILVFLDVMSEKFLSRLFLLSRS
      (IN REF. 1)
SQ SEQUENCE 747 AA; 60528 MW; 850E44B0D649E012 CRC64;

Query Match 68.4%; Score 1015; DB 1; Length 747;
Best Local Similarity 59.2%; Pred. No. 7.6e-44;
Matches 238; Conservative 8; Mismatches 14; Indels 142; Gaps 13;

QY 11 GSNASRGGLGGGAGAAAA-----AAAAAGAGGGGGY 45
DB 93 GSGCAGRGGLGGGAGAAAAAGAGAGGGGGLGNQAGRGGGGAGAAAAAGAGGGGGY 152
QY 46 GLGSGQTSRGGLGGGAGAAAAAGAGAGGGGGYGLGSGQTSRGGL 96
DB 153 GLGSGQ-AGRGGLGGGAGAAAAAGAGAGGGGGYGLGSGQ-AGRGGL 210
QY 97 GGGGAGAAAAA-----AAAAAGAGGGGGYGLGSGQTSRGGLG 135
DB 211 GGGGAGAAAAAGAGAGGGGGYGLGSGQAGAGAGGGGGYGLGSGQ-AGRG 266
QY 136 GCGAGAAAAA-----AGAGGGGGYGLGSGQTSRGGLGGGAGAAA- 182
DB 267 GEGAGAAAAAGAGAGGGGGYGLGSGQAGAGGGGGYGLGSGQ-AGRGGLGGGAGAGAG 325
QY 183 -----AAAAAGAGGGGGYGLGSGQTSRGGLGGGAGAAAA- 220
DB 326 QGGLGGGAGAGAGAAAAAGAGGGGGYGLGSGQ-AGRGGLGGGAGAGAGAGGAG 384
QY 221 -----AAAAAGAGGGGGYGLGSGQTSRGGLGGGAGAAAA- 259
DB 385 GYGGLGSGAGRGGGGAGAAAAAGAGAGGGGGYGLGSGQ-AGRGGLGGGAGAGAGAG 443
QY 260 -----AAAAAGAGGGGGYGLGSGQ 279
DB 444 GGAGGGGGYGLGSGAGRGGGGAGAAAAAGAGGGGGYGLGSGQ 485

RESULT 2
EBN1_EBV STANDARD; PRT; 641 AA.
AC F03211;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE EBNA-1 nuclear protein.
GN BKRF1.
OS Epstein-barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10377;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE-84270667; PubMed-6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tuffnell P.S., Barrell B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984).
RN [2]
RP SEQUENCE OF 1-26 FROM N.A.
RX MEDLINE-86259739; PubMed-3460083;
RA Sample J., Hummel M., Braun D., Birkenbach M., Kieff E.;
RT "Nucleotide sequences of mRNAs encoding Epstein-Barr virus nuclear
RT proteins: a probable transcriptional initiation site.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5096-5100(1986).
RN [3]
RP SUBCELLULAR LOCATION.
RX MEDLINE-90266473; PubMed-2161150;
RA Petti L., Sample C., Kieff E.;
RT "Subnuclear localization and phosphorylation of Epstein-Barr virus
RT latent infection nuclear proteins.";
RL Virology 176:563-574(1990).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 469-607.
RX MEDLINE-96006523; PubMed-753871;
RA Bochkarev A., Barwell J.A., Pfuetzner R.A., Furey W.F. Jr.,
RA Edwards A.M., Frappier L.;
RT "Crystal structure of the DNA-binding domain of the Epstein-Barr
RT virus origin-binding protein EBNA 1.";
RL Cell 83:39-46(1995).
CC -!- FUNCTION: INVOLVED IN LATENT CYCLE. EBNA-1 FUNCTIONS IN THE
CC MAINTENANCE REPLICATION OF EBV EPISOME. TRANSCRIPTIVATING FACTOR
CC FOR THE ORIGIN AND ENHANCER FUNCTIONS OF ORIP.
CC -!- SUBCELLULAR LOCATION: NUCLEAR. FREE IN THE NUCLEOPLASM, SOMEWHAT
CC ASSOCIATED WITH THE CHROMATIN AND HARDLY, IF AT ALL ASSOCIATED
CC WITH THE NUCLEAR MATRIX.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL; V01555; CAA24816.1; -.
DR EMBL; M13941; AAA45889.1; -.
DR PIR; A03773; Q0BE31.
DR PIR; S33021; S33021.
DR PDB; 1VH1; 23-DEC-96.
DR TRANSFAC; T00211; -.
DR InterPro; IPR004186; EBNA1.
DR Pfam; PF02905; EBNA1; 1.
KW Nuclear protein; DNA-binding; Transcription regulation; Activator;
KW 3D-structure.
FT DOMAIN 87 352 GLY/ALA-RICH.
SQ SEQUENCE 641 AA; 56427 MW; 4D161653E16FC341 CRC64;

Query Match 41.1%; Score 610.5; DB 1; Length 641;
Best Local Similarity 48.6%; Pred. No. 3.8e-24;
Matches 143; Conservative 14; Mismatches 104; Indels 33; Gaps 10;

QY 2 RGSHHHHHSGMASGRGLG-GQGAGAAAAAAGAGAGQ-GYGGLGSGQTSRGGLG 59
DB 83 KGTH-----GGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGA 130
QY 60 GQGAGAAAAAAGAGAGAGQ-GYGGLGSGQTSRGGLGGGAGAGAAAAAAG---GA 115
DB 131 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 188
QY 116 GGGYGLGSGQTSRGGLGGQ-GAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAG 174
DB 189 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 244
```

Qy	175	QGAGAAAAAAGGAGGAGGGYGLGSGQTSG-----RGLGGGAGAGAAAAAAGGAG	230
Db	245	AGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	304
Qy	231	QGGYGGGLGSGQTSRGGGLGGGAGAGAAAAAAGGAGGAGGGYGGGLGSGQTSGR	284
Db	305	AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	354
RESULT 3			
ID	Y208_MYCTU	STANDARD;	PRT; 1901 AA.
AC	OS3553;		
DT	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Hypothetical PE-PGRS family protein RV3508 precursor.		
GN	RV3508 OR RV023.15.		
OS	Mycobacterium tuberculosis.		
OC	Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;		
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.		
OC	NCBI_TaxID=1773;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=H37RV;		
RC	MEDLINE=98295987; PubMed=9634230;		
RA	COLE S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,		
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaita F.,		
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,		
RA	Davies R., Devlin K., Feldwell T., Gentles S., Hamlin N., Holroyd S.,		
RA	Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,		
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,		
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,		
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;		
RT	"Deciphering the biology of Mycobacterium tuberculosis from the		
RT	complete genome sequence.;"		
RL	Nature 393:537-544(1998).		
CC	-1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS		
CC	SUBFAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collabora-		
CC	tion between the Swiss Institute of Bioinformatics and the EMBL outstat-		
CC	ing the European Bioinformatics Institute. There are no restrictions on		
CC	use by non-profit institutions as long as its content is in no		
CC	modified and this statement is not removed. Usage by and for comm-		
CC	entities requires a license agreement (See http://www.isb-sib.ch/anno-		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; AL022022; CAAL7745.1; -		
DR	TubercuList; RV3508; -		
DR	InterPro: IPR000084; PE_region.		
DR	Pfam; PF00934; PE; 1.		
DR	ProDom; PD001223; PE_region; 1.		
KW	Hypothetical protein; Repeat; Signal; Complete proteome.		
FT	SIGNAL 1 30		
FT	CHAIN 31 1901		
FT	HYPOTHETICAL PE-PGRS FAMILY PROTEIN		
FT	RV3508.		
SQ	SEQUENCE 1901 AA; 147627 MW; C7B1923D5D0146CD CRC64;		
	Query Match 37.5%; Score 556.5; DB 1; Length 1901;		
	Best Local Similarity 44.9%; Pred. No. 3e-21;		
	Matches 140; Conservative 13; Mismatches 112; Indels 47; Gaps		
Qy	14 ASRGGLGGGAGAGAAAAAAGGAGGAGGGYGLGSGQTSGR-----GGLGGGAG 64		
Db	416 AGAGGGVGGGGNAIAGGINSGGAGGTGGQGGAGGAGGAGGAGGAGGAGGAGGAGG		
Qy	65 AAAAAAAGAGGAGGAGY-----GGLGSGQTSGRGGLGGGAGAGAAAAAAGGAGGAG-QGG 119		
Db	476 AGAGGAGCAAGTGGTGGVYCAAGKAGIGTGGGGAGGAGAGAGTATATCATGTTGFSGC 535		
Qy	120 YGGL----GSGQTSGRGGLGGGAGAAAAAAGGAGGAGQ-----GGYGGGLSGQTSGR 169		



```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z80225; CAB02341.1; -
DR EMBL; AE007103; AAK47026.1; ALT_INIT.
DR TIGR; MT2712; -.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE_region; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 51 51 V -> L (IN REF. 2).
FT CONFLICT 63 63 Q -> H (IN REF. 2).
FT CONFLICT 274 274 A -> T (IN REF. 2).
SQ SEQUENCE 778 AA; 63131 MW; DAB20FE58E4999E7 CRC64;

Query Match 34.88; Score 516.5; DB 1; Length 778;
Best Local Similarity 42.18; Pred. No. 1.6e-19;
Matches 136; Conservative 9; Mismatches 122; Indels 56; Gaps 16;

QY 14 ASGRGGLGGGGA--GAAAAAAGAGG-----AGGGVGGGLSGGTS-----GR 55
Db 454 AGGAGGAGGIGTDSAGGATGEGGAGTGVGGVGGAGGKGGGHNTGVGDAGG 513

QY 56 GGLGGGAGAAAAA--AAAGGAGGCGYGG--LGSQTSRGLGGGGAGAA--AAAA 109
Db 514 GGIGDGNALGAAGNGTGGAGNGGRCGMLGNAGGAGGTGCTGGGAGFAGGV 573

QY 110 AAAGAG-----QGGYGLGS-QTSSRGLGGGGA-----GAAAAAAGAGGAG 156
Db 574 GGAGGEGTLDGAGTAEGGTGGLGGLGGVGTGGMGGSGVGGNGAAGSLGLGGGGAG 633

QY 157 GYGLGSGQTSRGLGGQ---GAGAAAAAAGAGGAGGCGYGGVGGTSGRGLGG 212
Db 634 GVGTG--GIGGAGGAGNGAGGAGTTGGATIGGGGTGGVGGAG--GTGTGGAGG 689

QY 213 Q--GAGAAA-----AAAAAGGAGGCGYGGVGGTSGRGLGGGAGAAAAAAG 265
Db 690 TTGGCGCAGGLIGWAGAGGTGAGTGGCGGLGGQ--GNGNGGTGATGCGGGDFALGG 747

QY 266 GAGGGVGL--GSQTSGRIRP 286
Db 748 NGGAGGAGGSPGSGSIGQNGMP 770

RESULT 7
FBOH_BOMMO STANDARD; PRT; 5263 AA.
AC P05790; Q26379; Q17220;
DT 01-NOV-1988 (Rel. 09, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fibroin heavy chain precursor (Fib-H) (H-fibroin).
GN FIBH.
OS Bombyx mori (Silk moth).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC Ditrysia; Bombycoidea; Bombycidae; Bombyx.
OX NCBI_TaxID=7091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20330362; PubMed=10871375;
RA Zhou C.-Z., Confalonieri F., Medina N., Zivanovic Y., Esnault C.,
RA Yang T., Jacquet M., Janin J., Duguet M., Perasso R., Li Z.-G.;
RT "Fine organization of Bombyx mori fibroin heavy chain gene.";
RL Nucleic Acids Res. 28:2413-2419(2000).
RN [2]
RP SEQUENCE OF 1-168 FROM N.A.
RX MEDLINE=80045039; PubMed=498286;
```

```
RA Tsujimoto Y., Suzuki Y.;
RT "The DNA sequence of Bombyx mori fibroin gene including the 5'
RT flanking, mRNA coding, entire intervening and fibroin protein coding
RT regions.";
RL Cell 18:591-600(1979).
RN [3]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=79211211; PubMed=455439;
RA Tsujimoto Y., Suzuki Y.;
RT "Structural analysis of the fibroin gene at the 5' end and its
RT surrounding regions.";
RL Cell 16:425-436(1979).
RN [4]
RP PARTIAL SEQUENCE FROM N.A.
RX STRAIN=Kinshu X Showa;
RA Mita K., Ichimura S., Zama M., James T.C.;
RT "Specific codon usage pattern and its implications on the secondary
RT structure of silk fibroin mRNA.";
RL J. Mol. Biol. 203:917-925(1988).
RN [5]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=94365842; PubMed=7916056;
RA Mita K., Ichimura S., James T.C.;
RT "Highly repetitive structure and its organization of the silk fibroin
RT gene.";
RL J. Mol. Evol. 38:583-592(1994).
RN [6]
RP SEQUENCE OF 5179-5263 FROM N.A., AND DISULFIDE BONDS.
RX STRAIN=J-139;
RA Tanaka K., Kajiyama N., Ishikura K., Waga S., Kikuchi A., Ohtomo K.,
RA Takagi T., Mizuno S.;
RT "Determination of the site of disulfide linkage between heavy and
RT light chains of silk fibroin produced by Bombyx mori.";
RL Blochim. Biophys. Acta 1432:92-103(1999).
CC -1- FUNCTION: FORMS THE SILK FILAMENT; A STRONG, INEXTENSIBLE,
CC INSOLUBLE AND CHEMICALLY INERT FIBRE.
CC -1- SUBUNIT: FORMED OF TWO CHAINS: HEAVY AND LIGHT, THAT ARE LINKED BY
CC A DISULFIDE BOND. HEAVY-LIGHT CHAIN ASSEMBLY IS ESSENTIAL FOR THE
CC EFFICIENT INTRACELLULAR TRANSPORT AND SECRETION OF FIBROIN.
CC -1- TISSUE SPECIFICITY: PRODUCED EXCLUSIVELY IN THE POSTERIOR (PSG)
CC SECTION OF SILK GLANDS.
CC -1- DOMAIN: COMPOSED OF ANTIPARALLEL BETA SHEETS. THE STRANDS OF THE
CC BETA SHEETS RUN PARALLEL TO THE FIBER AXIS. LONG STRETCHES OF SILK
CC FIBROIN ARE COMPOSED OF MICROCRYSTALLINE ARRAYS OF (-GLY-SER-GLY-
CC ALA-GLY-ALA-)N INTERRUPTED BY REGIONS CONTAINING BULKIER RESIDUES.
CC THE FIBER IS COMPOSED OF MICROCRYSTALLINE ARRAYS ALTERNATING WITH
CC AMORPHOUS REGIONS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF226688; AAF76983.1; -
DR EMBL; V00094; CAA23432.1; -
DR EMBL; V00097; CAA23433.1; -
DR EMBL; S74439; AAB31861.1; -
DR EMBL; X13869; CAA32076.1; -
DR EMBL; M35378; AAA27839.1; -
DR EMBL; AB017362; BAA33147.1; -
DR PIR; S01844; S01844.
KW Silk; Signal; Repeat.
FT SIGNAL 1 21
FT CHAIN 22 5263
FT DOMAIN 149 5206
FT DISULFID 5244 5244
FT DISULFID 5260 5263
FT CONFLICT 10 10
C -> V (IN REF. 2).
POTENTIAL.
FIBROIN HEAVY CHAIN.
HIGHLY REPETITIVE.
INTERCHAIN (WITH LIGHT CHAIN).
```









Best Local Similarity 38.6%; Pred. No. 3.3e-18;  
Matches 133; Conservative 14; Mismatches 120; Indels 78; Gaps 14;

QY 11 GSMASGRGGLGGCGAGAAAAAAAAGA-----GQGGYGGLGSQGTSGRGLGGQG--- 62  
Db 425 GGWLFGNGNGGQTIGGVNCGAGGAGGAGILFTGTCGSGGPATGLGGICGAGAA 484

QY 63 -----AGAAAAAAAAGG-----AGQGGYGGLGSQGTSGRGLGGQGAGGA 103  
Db 485 LLFGSGGAGSGGACAVGGNGAGGNAGALLGAAGAGGAGGAGAVG--GNGGAGNG--GL 541

QY 104 AAAAAAAGG----ACQGGYGGLGSQ-GTSGRCGLGGCGAGAAAAAAAAGG----AG 154  
Db 542 FANGGAGGPGGFGSPAGAGGCGAGGCGGLFCAGGTGGAGCGGSTLAGGAGGAGNGGLFG 601

QY 155 QGGYGGLGSQGT-----GRGLGGQ-----GAGNAAAAAAAAAAGCA----- 191  
Db 602 AGCTGGAGSHSTAAGVSSGAGGAGGADAGLLSLGASGAGGSGGSSLTAAGVVGGIGGAGG 661

QY 192 ---GQGGYGGLGSQGTSGRGLGGQGAGGA-----AAAAAAAAGGAGQGGYGLGSQG 241  
Db 662 LLFGSGGAGSGGPFNSNGNGGAGGAGGDAGLLVGSGGAGGAGASATGAATGGDGAG--G 719

QY 242 TSGRGGLGGGAGAAAAAAAAGGACQGGYGG-----LGSGQTSG 282  
Db 720 KSGAFGLGGDGGAGGATGCSAFHIGGKGVGYSVAVLIGNGNGG 764

RESULT 13  
YI18\_MTCU

ID YI18\_MYCPU STANDARD; PRT; 498 AA.  
AC Q50615;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical PE-PGRS family protein RV1818c.  
GN RV1818C OR MT1866 OR MTCYAI1.25C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacterla (class); Actinobacteridae;  
OX Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RAD Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies K., Devlin K., Feitwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Horneby T., Jagels K., Krogh A., McLean L., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RL complete genome sequence."  
RL Nature 393:537-544 (1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Blais W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains."  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS  
SUBFAMILY.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at Basel.

ID	ELLS_MOUSE	STANDARD;	PRT;	860 AA.
DR	EMBL; Z78020; CAB01461.1; "			
DR	EMBL; AE007045; AAK46139.1; ALT_INIT.			
DR	TIGR; MT1866; "			
DR	TubercuList; Rv1818c; "			
DR	InterPro: IPR000084; PE_region.			
DR	Pfam; PF00934; PE; 1.			
DR	ProDom; PD001223; PE_region; 1.			
KW	Hypothetical protein; Transmembrane; Complete proteome.			
FT	TRANSMEM 1 21 POTENTIAL.			
FT	TRANSMEM 39 59 POTENTIAL.			
FT	TRANSMEM 93 113 POTENTIAL.			
FT	TRANSMEM 135 155 POTENTIAL.			
FT	TRANSMEM 158 178 POTENTIAL.			
FT	TRANSMEM 182 202 POTENTIAL.			
FT	TRANSMEM 209 229 POTENTIAL.			
FT	TRANSMEM 244 264 POTENTIAL.			
FT	TRANSMEM 281 301 POTENTIAL.			
FT	TRANSMEM 311 331 POTENTIAL.			
FT	TRANSMEM 338 358 POTENTIAL.			
FT	TRANSMEM 373 393 POTENTIAL.			
FT	TRANSMEM 403 423 POTENTIAL.			
FT	CONFLICT 183 212 MISSING (IN REF. 2).			
FT	CONFLICT 414 414 G -> GAGG (IN REF. 2).			
SO	SEQUENCE 498 AA; 40755 MW; 476F78F2482586BA CRC64;			
Query Match				
Best Local Similarity 32.9%; Score 489; DB 1; Length 498;				
Matches 132; Conservative 14; Mismatches 123; Indels 60; Gaps 14;				
Qy	11	GSMASGRGLGGQGA-----GAAAAAAGAGGAGGCGYGGGLSQGTS-----GRGGLG 59		
Db	173	GNVASTAGTGGAGGAGGGLLYGAGGAGGAGGRAGGCGVGGIGGAGGNGCLLFCAGGAG 232		
Qy	60	GQAGAAAAAAAANG-----ACGGYGGGLGSQGTS-----CRGGLG 98		
Db	233	GVG-GLAAGDGGAGGDDGLFFVGAGGAGGCTGTNTGGAGGAGNGGILLFCAGGVGG 291		
Qy	99	QGAGAAAAAAGAGGAGGCGYGL-LSQGTSGRGLG--GQAGAAAAAAGAGGAGGAG 155		
Db	292	VGGDVAFLEGTAPGPGGAGGAGGLFVGAGGAGGAGGLVNGGAGGSGGSLWGGGA 351		
Qy	156	GGYGLGSQGTSGRGLGGQ-----GACAAAAAAGAGGCGYGGGLSQGTS-----G 206		
Db	352	GCAGGVGST-TGGAGGAGGNAGLLVAGGAGGAGCAGCAGATGCGAGGNG--GTAGLLFG 408		
Qy	207	RGGLGGQAGAAAAA--AAAGAGGCGYGGGLGSQGTSGRGLGGGAGG-----255		
Db	409	AGGAGGFGCGAGGAGGLGRKAGLIGDGGDGCAGGNGCTGAKGDDGAGGGAILVNGNG 468		
Qy	256	AAAAAAGAGGAGGCGYGG-LGSQGTSGI 283		
Db	469	GNAGSGTPNGAGTGGAGGLGKNGMGL 497		
RESULT 14				
ELLS_MOUSE				
ID	ELLS_MOUSE	STANDARD;	PRT;	860 AA.
AC	P54320;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	Elastin precursor (Tropoelastin).			
GN	ELN.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.			
OX	NCBI TaxID=10090;			



[illegible]

Search completed: December 18, 2002, 16:17:54  
Job time : 9.46662 secs

***This Page Blank (uspto)***





```
QY 147 AAAAGAGGCGGGLGSGTSGRGLGGGCGAGAAAAA-----AGCAGGCGGGLG 200
DB 237 AAAAGAGGCGGGLGSG-----AGRGGCGAGAAAAAAGCAGGCGGCGGGLG 292
QY 201 SQTSGRGLGGGCGAGAAAAAAGGAGCGGGLG-----SQTSGRGLG-----G 250
DB 293 SQG-AGRGLGGGAG--AAAAAAGGAGGGLGGGAGAAAAAAGGAGGGLGGGAG 349
QY 251 QGAGAAAAAAGGAGCGGGLGSG 279
DB 350 QGAGAAAAAAGGVRGGGGLGSG 378

RESULT 4
Q9BIU1 PRELIMINARY; PRT; 447 AA.
AC Q9BIU1
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Major ampullate spidroin 1 (Fragment).
GN MASPI.
OS Argiope aurantia.
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneolidae; Araneidae; Argiope.
OX NCBI_TaxID=15684;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21179804; PubMed=11283372;
RA Gatesy J., Hayashi C., Motriuk D., Woods J., Lewis R.;
RT "Extreme Diversity, Conservation, and Convergence of Spider Silk
Fibroin Sequences";
RL Science 291:2603-2605(2001).
DR EMBL; AF350262; AAK30591.1; -.
DR PRINTS; PR000817; PRION.
FT NON_TER 1
FT NON_TER 447
SQ SEQUENCE 447 AA; 35630 MW; 78D160B7E6DF24C7 CRC64;

Query Match 55.1%; Score 818.5; DB 5; Length 447;
Best Local Similarity 60.2%; Pred. No. 2.5e-41;
Matches 197; Conservative 14; Mismatches 55; Indels 61; Gaps 13;

QY 15 SGRGLGGCGAGAAAAAAGCAGCGGGLGSGTSGRGLGGGCGAGAAAAA 74
DB 37 SGLGGGCGAGGAAAAAAGGCGGCGGGLGSG-AGCGGCGAGGAGAAAAA 95
QY 75 GGAGCGGCGGGLGSGT-----SGRGLGGGCGAGAAAAAAGGAGCGGGLGSGT 129
DB 96 GGAGGAGGGLGAGGAGCGGGLGCGGAGCGGAGAAAAAAGCGGCGGGLGSG- 153
QY 130 GRGLGGGCGAGAAAAAAGGAGCGGGLGSGTSGRGLGGGCGAGAAAAAAG 189
DB 154 --AGCGGAGRG-AAAAAAGGCGGGLGSG-AGCGGCGAGGAGAAAAAAG 209
QY 190 GAGCGGCG-----GLGSGTSGR-----GGLGGGCGAG----- 216
DB 210 GAGGGLGAGGAGCGGCGGCGGCGGAGCGGAGAAAAAAGCGGCGGCGAGCGGA 269
QY 217 --AAAAAAGGAGCGGGLGSGTSGRGLG-GGAGAAAAAAGGAGGAGCG- 270
DB 270 GRGAAAAAAGGCGGCGGGLGSG-AGCGGCGAGGAGAAAAAAGGAGGAGRG 328
QY 271 -----GYG-GLGSGTSGIRPAA 288
DB 329 LGAGGAGCGGGLGCGGAGGAGGAAS 355

RESULT 5
Q9BIU7 PRELIMINARY; PRT; 648 AA.
AC Q9BIU7
```

```
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Major ampullate spidroin 1 (Fragment).
GN MASPI.
OS Argiope trifasciata.
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneolidae; Araneidae; Argiope.
OX NCBI_TaxID=15684;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21179804; PubMed=11283372;
RA Gatesy J., Hayashi C., Motriuk D., Woods J., Lewis R.;
RT "Extreme Diversity, Conservation, and Convergence of Spider Silk
Fibroin Sequences";
RL Science 291:2603-2605(2001).
DR EMBL; AF350266; AAK30595.1; -.
DR HSSP; P10969; 1WGT.
FT NON_TER 1
FT NON_TER 648
SQ SEQUENCE 648 AA; 52887 MW; ACD867067B2302CD CRC64;

Query Match 55.1%; Score 817.5; DB 5; Length 648;
Best Local Similarity 51.8%; Pred. No. 3.8e-41;
Matches 200; Conservative 14; Mismatches 45; Indels 127; Gaps 14;

QY 15 SGRGLGGCGAGAAAAAAGCAGCGGGLGSGTSGRGLGGGCGAGAAAAA 73
DB 142 AGLGGGCGAGCGGAAAAAAGGCGGCGGGLGSG-AGCGGCGAGGAGAAAAA 199
QY 74 AGGAGCGGCGGGLGSG-OG-TSGRGLGGGCGAGAAAAAAGG-AGCGGCGGSGT-- 128
DB 200 AGGAGCGGCGGAGAGCGGCGGCGGAGGAAAAAAGGCGGCGGCGGCGG 259
QY 129 -----SGRGLGGCGA-----GAAAAAAG 151
DB 260 GYGCGGCGAAAAAAGGAGGCGGGLGAGCGGCGGCGGCGGCGGCGGCGG 319
QY 152 GAGCGGCGGGLGSGT-----SRRGLGGCGA-----SGRGLG 173
DB 320 QGCGGCGGGLGSGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 379
QY 174 GCGAGAAAAAAGG-AGCGGCGGGLGSGT-----SGRG-- 208
DB 380 GAGCGSASAAAAAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 439
QY 209 -----GLGCG-----AGAAAAAAGGAGCGGCGGCGGCGGCGGCGG 253
DB 440 GAGGAGCGGCGAGLGGCGGAGCGGAGAAAAAAGGCGGCGGCGGCGGCGG 495
QY 254 GAAAAAAGGAGCGGCGGGLGSG 279
DB 496 GGAAGGAGGCGGCGGCGGGLGSG 521

RESULT 6
Q9BIU3 PRELIMINARY; PRT; 691 AA.
AC Q9BIU3
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Fibroin 2 (Fragment).
OS Dolomedes tenebrosus.
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Lycosoidea; Pisauridae; Dolomedes.
OX NCBI_TaxID=15684;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21179804; PubMed=11283372;
RA Gatesy J., Hayashi C., Motriuk D., Woods J., Lewis R.;
RT "Extreme Diversity, Conservation, and Convergence of Spider Silk
Fibroin Sequences";
RL Science 291:2603-2605(2001).
```

```
DR EMBL; AF350270; AAK30599.1; -.
DR InterPro; IPR000104; Antifreeze_1.
DR PRINTS; PR002952; Eggshell.
DR PRINTS; PR00308; ANTIFREEZE1.
DR PRINTS; PR01228; EGGSHELL.
FT NON_TER 1
SQ SEQUENCE 691 AA; 51176 MW; CB46EF9F1991F33E CRC64;

Query Match
Best Local Similarity 51.8%; Score 768.5; DB 5; Length 691;
Matches 196; Conservative 17; Mismatches 51; Indels 115; Gaps 18;

QY 11 GSNASGRGLG-----GQAGAAAAAAG--GAGGGYGLGSGQT-----52
DB 173 GGAGSGGGYGGGGLGGYGGAGAGAAAAAGAGAGGGYGGGCGGAGAAATAT 232
QY 53 -----SGRGGGLGQ-----GAGAAAAAAG--GAGGGYGLGSGQT-----52
DB 233 AAAGAGSGGGYGGGGLGGYGGAGAGAAAAAGAGAGGGYGGG-----GQ 287
QY 94 GGLGGGAGAAAAAAGAGGGYGGGGLGSGQT-----SGRGGGLG 135
DB 288 GGY-GGAGAGAAAAAGAGAGGGYGGGAGAAAAAAGAGAGGGYGGG 346
QY 136 GQ-----GAGAAAAAAG--GAGGGYGLG-----SQ 164
DB 347 GGGLGGYGGGAGAGAAAAAGAGGGYGGGGLGGYGGAGAGAAAGAGSGGS 406
QY 165 GTSGRGLG--GGGAGAAAAAAGAGGGYGGG-----SGTSGRGLG 210
DB 407 GGAGGGGLGGYGGGAGAGAAAAAGAGGGYGGG-----GGL 464
QY 211 GGQGAGA-AAAAAAGAGGGYGG--GLGSGTSGRGLG--GQAGAAAAAAG 264
DB 465 GGVGGSGAGAAAGAGGGYGGGAGAGAGAGGGYGGGAGAGAGGGYGG 524
QY 265 G-GAGGGYGLGSGQTSG 282
DB 525 GAGAGGGYGGGGLGGY 543

RESULT 7
Q9BIU4 PRELIMINARY; PRT; 854 AA.
AC Q9BIU4;
AT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE Fibroin 1 (Fragment).
OS Dolomedes tenebrosus.
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Lycosoidea; Pisauridae; Dolomedes.
OX NCBI_TaxID=156846;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21179804; PubMed=11283372;
RA Gatesy J., Hayashi C., Motriuk D., Woods J., Lewis R.;
RT "Extreme Diversity, Conservation, and Convergence of Spider Silk
Fibroin Sequences."
SC Science 291:2603-2605(2001).
DR EMBL; AF350269; AAK30598.1; -.
DR InterPro; IPR000217; Tubulin.
DR PROSITE; PS00227; TUBULIN; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 854
SQ SEQUENCE 854 AA; 71881 MW; 5C52F1B0AE31C6C2 CRC64;

Query Match
Best Local Similarity 46.4%; Score 689.5; DB 5; Length 854;
Matches 197; Conservative 12; Mismatches 61; Indels 163; Gaps 23;

QY 11 GSNASGR-----GGLGGGAGAAAAAAGAGGGYGGGGLGSGQTSGRGLG--GQGA 63
DB 173 GGAGSGGGYGGGGLGGYGGAGAGAAAAAGAGAGGGYGGGCGGAGAAATAT 232
QY 53 -----SGRGGGLGQ-----GAGAAAAAAG--GAGGGYGLGSGQT-----52
DB 233 AAAGAGSGGGYGGGGLGGYGGAGAGAAAAAGAGAGGGYGGG-----GQ 287
QY 94 GGLGGGAGAAAAAAGAGGGYGGGGLGSGQT-----SGRGGGLG 135
DB 288 GGY-GGAGAGAAAAAGAGAGGGYGGGAGAAAAAAGAGAGGGYGGG 346
QY 136 GQ-----GAGAAAAAAG--GAGGGYGLG-----SQ 164
DB 347 GGGLGGYGGGAGAGAAAAAGAGGGYGGGGLGGYGGAGAGAAAGAGSGGS 406
QY 165 GTSGRGLG--GGGAGAAAAAAGAGGGYGGG-----SGTSGRGLG 210
DB 407 GGAGGGGLGGYGGGAGAGAAAAAGAGGGYGGG-----GGL 464
QY 211 GGQGAGA-AAAAAAGAGGGYGG--GLGSGTSGRGLG--GQAGAAAAAAG 264
DB 465 GGVGGSGAGAAAGAGGGYGGGAGAGAGAGGGYGGGAGAGAGGGYGG 524
QY 265 G-GAGGGYGLGSGQTSG 282
DB 525 GAGAGGGYGGGGLGGY 543

RESULT 8
Q9BIU0 PRELIMINARY; PRT; 360 AA.
AC Q9BIU0;
AT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE Major ampullate spidroin 1 (Fragment).
GN MASPI.
OS Latrodectus geometricus.
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneioidea; Theridiidae; Latrodectus.
OX NCBI_TaxID=156851;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21179804; PubMed=11283372;
RA Gatesy J., Hayashi C., Motriuk D., Woods J., Lewis R.;
RT "Extreme Diversity, Conservation, and Convergence of Spider Silk
Fibroin Sequences."
SC Science 291:2603-2605(2001).
DR EMBL; AF350273; AAK30602.1; -.
DR InterPro; IPR000104; Antifreeze_1.
DR InterPro; IPR002952; Eggshell.
DR PRINTS; PR00308; ANTIFREEZE1.
DR PRINTS; PR01228; EGGSHELL.
DR PRINTS; PR00341; PRION.
FT NON_TER 1
FT NON_TER 360
SQ SEQUENCE 360 AA; 30253 MW; 0EB6E46EC1122740 CRC64;

Query Match
Best Local Similarity 45.8%; Score 680; DB 5; Length 360;
Matches 176; Conservative 12; Mismatches 44; Indels 75; Gaps 17;

QY 15 SGRGGGLG-----GQAGAAAAAAGAGGGYGGGGLGSGQTSGRGGYGGGAGAA 66
DB 3 SQGGGGYGGYGGGAGAGAGAGAAAAAAGAGGGYGGGCGGAGAAATAT 232
QY 67 AAAAAAGAGGGYGG-----GLGSGTSGRGGYGGGCGGAGAGAA 107
DB 58 GAAAAAGAGGGYGGGAGAGAGAAAAAAGAGGGYGGGCGGAGAGAA 117
```





```
FT NON_TER 1 1
FT NON_TER 253
SQ SEQUENCE 253 AA: 21311 MW; 9D6B1D34A69B9869 CRC64;

Query Match 42.7%; Score 634; DB 5; Length 253;
Best Local Similarity 61.7%; Pred. No. 1.1e-30;
Matches 150; Conservative 5; Mismatches 12; Indels 76; Gaps 10;

QY 19 GLGCGGAGAAAAAAGAGGCGGCGGLGSGTSGRGLGGGCGAGAAAAAAGAGGAG 78
D 1 GLGCGGAGGAGAAAAAAGAGGCGG-----GGGCGGAG-----AAAAGGAG 45

QY 79 QGCGGLGSGTSGRGLGGGCGAGAAAAAAGAGGCGGCGGLGSGTSGRGLGGG 138
D 46 QGCG-----GLGCGR-----AAAGGAGCGG-----GGLGGG 74

QY 139 AGAAAAAAGAGGCGGCGGLGSGTSGRGLGGGCGAGAAAAAAGAGGAGCGG 198
D 75 AGGAGAGAAAAAGAGGCGG-----GGGCGGAG--AAAAAAGGAGGCGG 121

QY 199 LGSQGTSGRGLGGGCGAGAAAAAAGAGGCGGCGGLGSGTSGRGLGGGCGAAAA 258
D 122 LGSQG-AGRGYGGCGAG-----AAVAAIGVGCGGCGGVS-----GASAASA 164

QY 259 AAA 261
D 165 AAS 167

RESULT 12
Q9BIT9 PRELIMINARY; PRT; 373 AA.
AC Q9BIT9;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Major ampullate spideroin 2 (Fragment).
GN MAS2.
OS Latrodectus geometricus.
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneolidae; Theridiidae; Latrodectus.
OX NCBI_TaxID=156851;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21179804; PubMed=11283372;
RA Gately J., Hayashi C., Motriuk D., Woods J., Lewis R.;
RT "Extreme Diversity, Conservation, and Convergence of Spider Silk
  Fibroin Sequences."
RL Science 291:2603-2605(2001).
DR EMBL; AF350274; AAK30603.1; -.
DR InterPro; IPR000104; Antifreeze_1.
DR PRINTS; PR00308; ANTIFREEZE1.
DR PRINTS; PR01574; TUBBYPROTEIN.
FT NON_TER 1 1
FT NON_TER 373 373
SQ SEQUENCE 373 AA: 29724 MW; ED8E0E2E16CA59E8 CRC64;

Query Match 42.6%; Score 632.5; DB 5; Length 373;
Best Local Similarity 52.1%; Pred. No. 1.8e-30;
Matches 170; Conservative 17; Mismatches 68; Indels 71; Gaps 17;

QY 14 ASGRGLG-----CGGAGAAAAAAGAGGCGGCGGLGSGTSGRGLGG 60
D 1 ASASGGAGGPGRQCGYGGGSGASAAAAAAGGAGGPGGCGGPGSGPGAGCGG 60

QY 61 QGAGAAAAAAGAGGCGG-----GLGSGTSGRGLGGGCGAGAAA----- 105
D 61 QG-GSGASAAAAAAGSGGPGGPGGAGGPGSGPGSGGAAAAAASGPGGCGG 119

QY 106 -----AAAAAAGGAGCGGCG-GLGSGTSGR-----GLGSGGAGAAAAAAGGA 153
D 120 GPGASAAAAAAGSGGPGGCGGPGSGYGGPGAGCGGCGGCG-GSAAAAAAGS 178
```

```
QY 154 GCGG-----GLGSGTSGRGLGGGCGAGAAAAAAGAGGCGG-----GGLGSGTSGR 207
D 179 GRGCGPGGAACPCNYGPGSGG-----AAASAAAAAGGCGGCGGCGGAAAAAAS 232

QY 208 GGLG-----GCGAGAAAAAAGAGGCGGCGGLGSGTSGRGLGGGCGAGAAAAA 260
D 233 GGAAGPQCGGCGGCGGAGAAAAAAGAGGCGGCG-----QCPAGYGP-GCGG-GSGGAA 286

QY 261 AAAAGGAGCGG-----GLGSGTSG 282
D 287 AAAASGPGCGYCPGAGPCNYGPG 312

RESULT 13
O76786 PRELIMINARY; PRT; 2639 AA.
AC O76786;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Fibroin.
OS Antheraea pernyi (Chinese oak silk moth).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Bombycoidea; Saturniidae; Saturniinae; Saturniini; Antheraea.
OX NCBI_TaxID=71119;
RN [1]
RP SEQUENCE 2639 AA; 216057 MW; 2EE3310DEEB09B9A CRC64;
RX MEDLINE=20496953; PubMed=11040284;
RA Sezutsu H., Yukuhiro K.;
RT "Dynamic rearrangement within the antheraea pernyi silk fibroin gene
  is associated with four types of repetitive units."
RL J. Mol. Evol. 51:329-338(2000).
DR EMBL; AF083334; AAC32606.1; -.
SQ SEQUENCE 2639 AA; 216057 MW; 2EE3310DEEB09B9A CRC64;

Query Match 41.6%; Score 617.5; DB 5; Length 2639;
Best Local Similarity 39.8%; Pred. No. 6.2e-29;
Matches 174; Conservative 10; Mismatches 90; Indels 163; Gaps 13;

QY 14 ASRGRGLG-----QGAGAAAAAAGAGGCGGCGGLGSGT----- 52
D 1142 ASGAGSGGGYGGWGDGSDSAAAAAAGAGGSGGGYGGYGGSDSAAAAA 1201

QY 53 -----SGRGLG-----QGAGAAAAAAGAGGCGGCGGLGSGT----- 90
D 1202 AAAAGSGAGGCGGCGWGDGSDSAAAAAAGAGGSGGGYGGYGGYGGSDS 1261

QY 91 -----SGRGLG-----QGAGAAAAAAGAGGCGGCGGLGSGT----- 121
D 1262 AAAAAAAGAGGAGGCGGCGWGDGSDSAAAAAAGAGGSGGGYGGYGGYGG 1321

QY 122 -----GLGSGTSGRGLG----- 136
D 1322 GSSAAAAAAGAGGCGGCGWGDGSDSAAAAAAGAGGSGGGYGGWGDGSD 1381

QY 137 --QGAGAAAAAAGAGGCGGCGGLGSGT-----SGRGLG----- 174
D 1382 SAAAAAAGAGGCGGCGGCGWGDGSDSAAAAAAGAGGSGGGYGGYGGWGDG 1441

QY 175 -----QGAGAAAAAAGAGGCGGCGGLG-----QGTSRGLG 212
D 1442 GYGSAAAAAAGAGGCGGCGGCGWGDGSDSAAAAAAGAGGAGGAGGCGG 1501

QY 213 QGAGAAAAAAGAGGCGGCGGCGWGDGSDSAAAAAAGAGGAGGAGGAGGAGG 271
D 1502 YGSDSAAAAAAGAGGCGGCGGCGWGDGSDSAAAAAAGAGGAGGAGGAGG 1561

QY 272 YGGLGSGTSGIRPAA 288
D 1562 DGCGS-GSAAAAA 1577
```



***This Page Blank (uspto)***

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:05:10 : Search time 45,4697 Seconds

(without alignments)  
1995.696 Million cell updates/sec

Title: US-09-490-291-4

Perfect score: 3493

Sequence: 1 MRGSHHHHHHGMASRGGL.....GGYGPQQQTSGIRRAAKLN 681

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3493	100.0	681	22 AAB82609	Spider recombinant
2	3390	97.1	691	22 AAB82610	Spider recombinant
3	2419	69.3	1255	23 AAM50037	N. clavipes spidro
4	2419	69.3	1880	23 AAM50039	N. clavipes spidro
5	2372	67.9	989	23 AAM50038	N. clavipes spidro
6	2294	65.7	1617	23 AAU11794	Drigline protein 1
7	2294	65.7	1626	23 AAU11798	Drigline protein 1
8	2254	64.5	818	23 AAU11797	Drigline protein 1
9	2250	64.4	809	23 AAU11793	Drigline protein 1
10	2139	61.2	630	23 AAM50042	N. clavipes spidro

11	2139	61.2	676	23 AAM50047	N. clavipes spidro
12	2135	61.1	651	20 AA40097	Spider silk protei
13	2135	61.1	651	23 AAU11781	Spider natural sil
14	2135	61.1	718	19 AAW53346	Nephila clavipes s
15	2135	61.1	718	21 AAW59070	N. clavipes spider
16	2126	60.9	718	12 AAR14308	N. clavipes draglin
17	2076	59.4	606	16 AAR99053	Spider dragline va
18	2076	59.4	606	20 AA40100	Polymer of an anal
19	2052.5	58.8	646	18 AAW27178	Nephila clavipes s
20	1972.5	56.5	606	16 AAR99055	Spider dragline va
21	1972.5	56.5	606	20 AA40102	Polymer of an anal
22	1971.5	56.4	604	16 AAR99057	Spider dragline va
23	1963.5	56.2	606	20 AA40101	Polymer of an anal
24	1711.5	49.0	528	22 AAB82611	Spider recombinant
25	1609.5	46.1	773	23 AAM50045	N. clavipes spidro
26	1609.5	46.1	777	23 AAM50046	N. clavipes spidro
27	1458.5	41.8	364	23 AAM50043	N. clavipes spidro
28	1425	40.8	291	16 AAB82608	Spider recombinant
29	1405.5	40.2	831	16 AAR80168	pMIS1 Misp spider
30	1395	39.9	615	20 AA40099	Spider silk protei
31	1387	39.7	655	23 AAB76672	Protein related to
32	1331.5	38.1	714	16 AAR99059	Spider dragline va
33	1331.5	38.1	714	20 AA40103	Polymer of an anal
34	1233.5	35.3	1177	9 AAP80940	SLPIII protein com
35	1233.5	35.3	1177	11 AAR05307	SLP III (Silk-fibr
36	1233.5	35.3	1177	17 AAR95105	Silk like protein
37	1233.5	35.3	1177	18 AAW26342	Silk-like protein
38	1233.5	35.3	1177	21 AAB78277	SLPIII amino acid
39	1233.5	35.3	1177	22 AAB72725	Repetitive protein
40	1233.5	35.3	1177	14 AAB63995	SLPIII protein seq
41	1233.5	35.3	1178	14 AAR41007	Silk-like protein
42	1233.5	35.3	1178	19 AAW53518	Amino acid sequenc
43	1176.5	33.7	1023	19 AAW53524	Amino acid sequenc
44	1176.5	33.7	1059	14 AAR41013	SLP4 multimeric pr
45	1176.5	33.7	1059	18 AAW26348	SLP4 synthetic pro

## ALIGNMENTS

RESULT 1  
AAB82609

ID AAB82609 standard; Protein: 681 AA.

XX AAB82609;

XX 02-OCT-2001 (first entry)

DT Spider recombinant silk protein PQE((SP1)4/(SP2)1)4.

DE Spider; orb-weaver; silk protein; PQE((SP1)4/(SP2)1)4;  
KW structural protein; purification; fibre; spinning.

XX Nephila clavipes.

OS WO200153333-A1.

PN 26-JUL-2001.

PD 01-NOV-2000; 2000WO-US30086.

XX 20-JAN-2000; 2000US-0490291.

XX (MELL/) MELLO C M.

PA (ARCI/) ARCIDIAcono S.

PA (BUTL/) BUTLER M M.

XX (USSA ) US SEC OF ARMY.

PI Mello CM, Arcidiacono S, Butler MM;

XX WPI; 2001-483136/52.

DR N-PSDB; AAW26302.

XX













QY 632 GAAAAAAGAGAGCGGGLGSGQTSRGGYGP-GQOTSG 672  
 Db 791 G-----AAAAAGAGCGGGLGSGQ-AGCGYGGGLGSGQAG 827

# RESULT 7

AAU11798  
 ID AAU11798 standard; Protein; 1626 AA.

XX AAU11798;

XX 26-MAR-2002 (first entry)

XX Dragline protein 1 analogue DP-1B/his tag 16mer.

XX Silk-like protein; SLP; transgenic plant; promoter; 5' terminator;  
 KW fabric production; material construction; rope; surgical suture;  
 KW flexible tie down; electrical component; implantation;  
 KW Dragline protein 1; DP-1B 16mer; His tag.

XX Nephila clavipes.

OS Synthetic.

XX WO200190389-A2.

PN 29-NOV-2001.

XX 24-MAY-2001; 2001WO-US16937.

XX 25-MAY-2000; 2000US-206968P.

XX (DUPO ) DU PONT DE NEMOURS & CO E. I.

XX Yang JG;

XX WPI; 2002-106209/14.

XX Producing silk-like proteins in a green plant, useful in fabrics and  
 PT material construction, comprises providing a silk-like protein  
 PT expression cassette to a green plant -

XX Example 1; Page 87-92; 93pp; English.

XX The invention relates to a method of producing silk-like proteins (SLP)  
 CC in a green plant by providing a green plant containing an SLP expression  
 CC cassette. The green plant contains the expression cassette P-SLP-T where  
 CC P is a promoter for driving the expression of an SLP transgene and T is a  
 CC 5' terminator. The method is useful for producing silks and silk-like  
 CC proteins in green plants. The silks and silk-like proteins may be used in  
 CC fabrics or in material construction, such as rope, surgical sutures,  
 CC flexible tie downs for certain electrical components, or as a  
 CC biomaterial for implantation. The method allows for more cost effective  
 CC production of silk not obtained from natural or microbial sources.  
 CC The present sequence is the silk protein dragline protein (DP) 1  
 CC synthetic variant DP1B/his tag 16mer (i.e. 16 copies of DP1B with a His  
 CC tag at the C-terminus) which is used as an SLP construct in the method of  
 CC the invention.

XX Sequence 1626 AA;

Query Match 65.7%; Score 2294; DB 23; Length 1626;  
 Best Local Similarity 61.6%; Pred. No. 7.1e-160;  
 Matches 543; Conservative 13; Mismatches 38; Indels 288; Gaps 36;

QY 11 GSMASGRGLGGGAGAGAAAA-----AAAAAGAGCGGGLGSG 49

Db 16 GSQAGRGRLGGGAGAGAAAAAGGAGCGGGLGSGQAGAGAGAGAGCGGGLGSG 75

QY 50 QGTSGRGLGGGAGAGAAAAAGGAGCGGGLGSGQGT-----SGRGLG 97

Db 76 QG-----AGRGCGQAG-----AAAAAGAGCGGGLGSGQAGCGGGLGSG 127

QY 98 GQAGAGAAAA-----AAAAAGAGCGGGLGSGQTSRGGGLG 136

Db 128 GQAGAGAGAGAGCGGGLGSGQAGAGAGAGAGAGAGCGGGLGSGQ-----ACRG 183  
 QY 137 QGAGAGAGAGAGAGAGCGGGLGSGQTSRGGYGP-GQOTSGRGLGGGAGAGAGAG- 194  
 Db 184 QGAG-----AAAAAGAGAGCGGGLGSGQ-AGCGYGGGLGSGQAGRGLGGGAGAGAG 238  
 QY 195 -----AAAAAGAGAGCGGGLGSGQTSRGGYGP-GQOTSGRGLGGGAGAGAGAG 234  
 Db 239 AAGGAGCGGGLGSGQAGAGAGAGAGAGAGCGGGLGSGQ-----AGRGCGAG- 290  
 QY 235 AAAAGGAGCGGGLGSGQGT-----SGRGLGGGAGAGAGAGAGAGAGAGAGAGAG 270  
 Db 291 AAAAGGAGCGGGLGSGQAGCGGGLGSGQAGCGGGLGSGQAGAGAGAGAGAGAGAGAG 350  
 QY 271 -----AAAAAGAGCGGGLGSGQTSRGGYGP-GQOTSGRGLGGGAGAGAGAGAGAGAG 321  
 Db 351 SQAGCGAGAGAGAGAGAGCGGGLGSGQ-----AGRGCGAG-----AAAAAGAGAGAG 402  
 QY 322 YGGLGSGQTSRGGYGP-GQOTSGRGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 359  
 Db 403 YGGLGSGQ-AGCGYGGGLGSGQAGRGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 461  
 QY 360 AAAAGGAGCGGGLGSGQTSRGGYGP-GQOTSGRGLGGGAGAGAGAGAGAGAGAGAGAGAG 418  
 Db 462 AAAAGGAGCGGGLGSGQ-----AGRGCGAG-----AAAAAGAGAGCGGGLGSGQAG 513  
 QY 419 -----SGRGLGGGAG 446  
 Db 514 QGGLGSGQAGRGLGGGAG 573  
 QY 447 YGGLGSGQTSRGGYGP-GQOTSGRGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 505  
 Db 574 YGGLGSGQ-----AGRGCGAG-----AAAAAGAGAGCGGGLGSGQ-AGCGYGGGLGSG 624  
 QY 506 TSGRGLGGGAG 544  
 Db 625 GAGRGLGGGAG 683  
 QY 545 SGRGLGGGAG 592  
 Db 684 ---AGRGCGAG-----AAAAAGAGAGCGGGLGSGQAGCGGGLGSGQAGRGLGGGAG 736  
 QY 593 AG 631  
 Db 737 AG 792  
 QY 632 GAAAAAG 672  
 Db 793 G-----AAAAAGAGAGCGGGLGSGQ-AGCGYGGGLGSGQAG 829

## RESULT 8

AAU11797

ID AAU11797 standard; Protein; 818 AA.

XX AAU11797;

XX 26-MAR-2002 (first entry)

XX Dragline protein 1 analogue DP-1B/his tag 8mer.

XX Silk-like protein; SLP; transgenic plant; promoter; 5' terminator;  
 KW fabric production; material construction; rope; surgical suture;  
 KW flexible tie down; electrical component; implantation;  
 KW Dragline protein 1; DP-1B 8mer; His tag.

XX Nephila clavipes.

OS Synthetic.

XX WO200190389-A2.

XX 29-NOV-2001.







Db	108	G	-----AAAAAAGGAGGGYGGGCGQ-----	AGRGQG	-----AAAAA142
Qy	200	AGGAGCGYGGGGLGSQGTSGRGGGCGGAGAAAAA	-----AGSAGGGYGGGCG	250	
Db	143	AGGAGCGYGGGGLGSQ-----AGRGGLGGGAGAAAAA	AGGAGCGYGGGCGGAGGGYGGGCG	201	
Qy	251	SQGTSGRGGGCGGAGAAAAA	-----AAAAGGAGCGYGGGCGS	289	
Db	202	SQG-----AGRGGLGGCGAGAAAAAAGGAGGGGCGGAGAGASAAAAGGAGGGYGGGCGS	260		

Db

261 QG----AGRGGAGAC-----AAAAAAGAGGGGGGLGGQG-AGGGGGYGGJGSQGAAGGGL 311

QY

290 QGTSGRRGGJGGGGAGAAAAAAAAGAGAGCGGTGGTGGTGCTSGPGGYGP-GGTTSGKGGGL 340

Db 312 GGQAGAAAAGGAGGGGGLGGQAGAGAGAAAAAAGAGGGGGGLSQG - AGRGGLGGQG 370

QY  
39T AGAAGGAGGGGGGGGCTSGKGGGGCGACAAAAAAAAGGAGGGGCGG 430

537 5' AG --- AAAAAAAAAAGGAGGGGGGGGGG --- AAGGGGGAG --- AAAAAAAAAAGGAGGGGGG 411  
 QY 451 LGSQGTSGRGGGLGGGAGAGAAAAAAGGAGGGGGGSGGT-----SGP 497  
 ||:||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Db 420 LGNQG-AGRGGGQGGAG---AAAAAAGGAGQGQGGYGGGLGNQCGAGRGGCGAAAAACGGAGQ 475

QY 498 GGYGP-GQQTSGRGGGAGAGAAAAA-----AGGAGQGGYGGGCGTSGRGG 550

470 GGTGGTGGAGR---GGQGAGAAAAAAGAGQEGIRGGAGQGGYGGTGSQG-SGRGGT 531

QY 331 GGQGAGAAA-----AAAGGAGGGYGGGTSQGTSGRGGGG 590

DD GGQGAGAAAAAGGAGTGGTGGAGQGAGAAAAAAGGVKQGGTGGTGSQG----AGRGG 58/

QY J31 Q8AGAAAAAAGGAGGGIGGLSSQISGRGGLGGQGAGHAAAAAGGAGGGI 050

[illegible]

**Q**

0  
0  
0  
—  
:  
—  
—  
<sup>4</sup>  
0  
0

2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 104

RESULT 13  
AAU11781

ID	AAU11781 standard; protein: 651 AA.
XX	
XX	AAU11781;
AC	
XX	
XX	26-MAR-2002 (first entry)
DT	
XX	
XX	Spider natural silk protein Spidroin 1.
DE	
XX	
XX	Orb-weaving spider; spidroin 1; silk-like protein; SLP; transgenic pl
KW	promoter; 5' terminator; fabric production; material construction; ro
KW	surical suture; flexible tie down; electrical component; implantation
KW	

XX Nephila clavipes.  
OS  
XX  
XX  
PN WO200190389-A2.  
XX  
PD 29-NOV-2001.

XX	24-MAY-2001; 2001WO-US16937.
PF	
XX	25-MAY-2000; 2000US-206968P.
PR	

XX  
PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
XX  
PI Yang JG;

XX  
DR WPI; 2002-106209/14.

XX Producing silk-like proteins in a green plant, useful in fabrics and  
PT material construction, comprises providing a silk-like protein  
PT expression cassette to a green plant -  
XX  
XX  
XX Disclosure; Page 69-71; 93pp; English.  
XX  
CC The invention relates to a method of producing silk-like proteins (SLP)  
CC in a green plant by providing a green plant containing an SLP expression  
CC cassette. The green plant contains the expression cassette P-SLP-T where  
CC P is a promoter for driving the expression of an SLP transgene and T is a  
CC 5' terminator. The method is useful for producing silks and silk-like  
CC proteins in green plants. The silks and silk-like proteins may be used in  
CC fabrics or in material construction, such as rope, surgical sutures,  
CC flexible tie downs for certain electrical components, or as a  
CC biomaterial for implantation. The method allows for more cost effective  
CC production of silk not obtained from natural or microbial sources.  
CC The present sequence is orb-weaving spider spidroin1, a natural silk  
CC protein whose sequence is used as the basis for the SLP proteins used in  
CC the method of the invention.  
XX  
XX Sequence 651 AA;

Query Match 61.1%; Score 2135; DB 23; Length 651;  
Best Local Similarity 68.3%; Pred. No. 1.4e-148;  
Matches 495; Conservative 15; Mismatches 39; Indels 176; Gaps 34;

QY 23 QGAGAAAAAAGAGGAGGGGGLGSGQT--SGRGLGGGAG-AAAAAAAAAGGAGQ 79  
DB 1 QGAG----AAAAAGAGGGGGLGGAGGGGGLGGAGGAGAGAAAAAGAGGQ 56  
QY 80 GGYGGLSGQTSRGLGGGAGAAAAAAGAGGAGGGGGLGSGQTSRGLGGGQ 139  
DB 57 GGYGGLSGQ----AGRGGGAG----AAAAAGGAGGGGGLGSGQ-AGRGLGGQ 107  
QY 140 GAAAAAAGAGGAGGGGGLGSGQTSRGLGGGAGAAAAAAGAGGAGGAG 199  
DB 108 G----AAAAAGGAGGGGGLGSGQ-----AGRGGG-----AAAA 142  
QY 200 AGGAGGGGGLGSGQTSRGLGGGAGAAAAAAGAGGAGGGGGLG 250

QY 201 AGGAGGGGGLGSGQTSRGLGGGAGAAAAAAGAGGAGGGGGLG 201  
QY 251 SGTSGRGLGGGAGAAAAA-----AAAAAGGAGGGGGLG 289  
DB 202 SGTSGRGLGGGAGAAAAAAGAGGAGGGGGLGSGQTSRGLGGG 260  
QY 290 QGTSGRGLGGGAGAAAAAAGAGGAGGGGGLGSGQTSRGLGGG 348  
DB 261 QG----AGRGGGAG----AAAAAGGAGGGGGLGSGQ-AGRGGLGG 311  
QY 349 GGGGAGAAAA-----AAAAAGGAGGGGGLGSGQTSRGLGGG 390  
DB 312 GGGGAGAAAAAGGAGGGGGLGGAGGAGAAAAAAGAGGAGGGG 370  
QY 391 AGAAAAAAGAGGAGGGGGLGSGQTSRGLGGGAGAAAAAAGAGGAGGG 450  
DB 371 AG----AVAAAAAGGAGGGGGLGSGQ-----AGRGGGAG----AAAAAGGAGGG 419  
QY 451 LGSQTSRGLGGGAGAAAAAAGAGGAGGGGGLGSGQTSRGLGGG 497  
DB 420 LGNQ-AGRGLGGGAG----AAAAAAGGAGGGGGLGSGQTSRGLGGG 475  
QY 498 GGYGP-GQGTSGRGLGGGAGAAAAA-----AGGAGGGGGLGSGQTSRGLGG 550  
DB 476 GGYGGLSGQAGR---GGGAGAAAAAAGAGGAGGGGGLGSGQ-AGRGGL 531  
QY 551 GGGGAGAAAA-----AAAAAGGAGGGGGLGSGQTSRGLGGG 590  
DB 532 GGGGAGAAAAAGGAGGGGGLGGGAGGAGAAAAAAGGVRGGGGLGSGQ----AGRG 587  
QY 591 QGAGAAAAAAGGAGGGGGLGSGQTSRGLGGGAGAAAAAAGAGGAGGG 650

DB 588 QGAG----AAAAAGGAGGAGGGGGLGGGQV-GRGGLGGGAG-----AAAAAGGAGGG 636  
QY 651 GGLGS 655  
DB 637 GGVGS 641  
RESULT 14  
AAW53346  
ID AAW53346 standard; Protein; 718 AA.  
XX  
AC AAW53346;  
XX  
DT 06-JUL-1998 (first entry)  
XX  
DE Nephila clavipes spider silk protein.  
XX  
KW Spider; Nephila clavipes; silk protein; tandem repeat; fibre; dragline;  
KW cocoon; tensile strength; elasticity.  
XX  
OS Nephila clavipes.  
XX  
PN US5728810-A.  
XX  
PD 17-MAR-1998.  
XX  
PF 19-APR-1995; 95US-0425069.  
XX  
PR 15-APR-1991; 91US-0684819.  
PR 20-APR-1990; 90US-0511792.  
PR 04-OCT-1994; 94US-0317844.  
PR 19-APR-1995; 95US-0425069.  
XX  
PA (UYWY-) UNIV WYOMING.  
XX  
PI Hinman MB, Lewis RV, Xu M;  
XX  
DR WPI; 1998-270437/24.  
DR N-PSDB; AAV23249.  
XX  
PT Recombinant spider silk proteins - useful for making fibres  
XX  
PS Claim 1; Column 29-34; 68pp; English.  
XX  
CC The present sequence represents a spider silk protein from the present  
CC invention. Spider silk proteins, and peptide fragments within the  
CC proteins, can be produced and purified independently and can then be  
CC mixed and made into fibres that have higher tensile strengths and  
CC elasticity than naturally occurring fibres. The fibres can be used in  
CC mixed composites. The invention allows the two naturally occurring  
CC Nephila clavipes silk proteins to be produced independently so that  
CC they can later be combined to form silk fibres of high tensile strength  
CC and elasticity.  
XX  
SQ Sequence 718 AA;

Query Match 61.1%; Score 2135; DB 19; Length 718;  
Best Local Similarity 68.3%; Pred. No. 1.5e-148;  
Matches 495; Conservative 15; Mismatches 39; Indels 176; Gaps 34;

QY 23 QGAGAAAAAAGAGGAGGGGGLGSGQT--SGRGLGGGAG-AAAAAAAAAGGAGQ 79  
DB 1 QGAG----AAAAAGGAGGGGGLGGAGGGGGLGGAGGAGAGAAAAAGGAGQ 56  
QY 80 GGYGGLSGQTSRGLGGGAGAAAAAAGAGGAGGGGGLGSGQTSRGLGGGQ 139  
DB 57 GGYGGLSGQ----AGRGGGAG----AAAAAGGAGGGGGLGSGQ-AGRGLGGQ 107  
QY 140 GAAAAAAGAGGAGGGGGLGSGQTSRGLGGGAGAAAAAAGAGGAGGAG 199  
DB 108 G----AAAAAGGAGGGGGLGSGQ-----AGRGGG-----AAAA 142  
QY 200 AGGAGGGGGLGSGQTSRGLGGGAGAAAAAAGAGGAGGGGGLG 250





***This Page Blank (uspto)***





STREET: 301 No. 5728810th Washington Street  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22046  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
APPLICATION NUMBER: US/08/425,069  
FILING DATE: 19-APR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 1447-106P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 718 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-425-069-2

Query Match 61.1%; Score 2135; DB 1; Length 718;  
Best Local Similarity 68.3%; Pred. No. 7.8e-151;  
Matches 495; Conservative 15; Mismatches 39; Indels 176; Gaps 34;

QY 23 QGAGAAAAAAGAGGAGGCGGGLGSGGT--SGRGLGGGAG--AAAAAAGAGGAG 79  
DB 1 QGAG-----AAAAAGGAGGCGGGLGGGAGGAGGAGGAGGAGGAGGAGGAG 56  
QY 80 GYGGLGSGTSGRGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 139  
DB 57 GYGGLGSGT--AGRGGAG-----AAAAAGGAGGCGGGLGSGGT--AGRG 107  
QY 140 GAAAAAAGAGGAGGCGGGLGSGGTSGPGYGPQGTSGRGLGGGAGGAGGAGGAG 199  
DB 108 G-----AAAAAGGAGGCGGGLGSGT--AGRGGAG-----AAAA 142  
QY 200 AGGAGGCGGGLGSGTSGRGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 250  
DB 143 AGGAGGCGGGLGSGT--AGRGGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 201  
QY 251 SGTSGRGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 289  
DB 202 SQG--AGRGGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 260  
QY 290 QGTSGRGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 348  
DB 261 QG-----AGRGGAG-----AAAAAGGAGGCGGGLGSGT--AGRG 311  
QY 349 GGCGAGAAAA-----AAAAAGGAGGCGGGLGSGTSGRGLGGGAGGAGGAG 390  
DB 312 GGCGAGAAAAAGGAGGCGGGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 370  
QY 391 AGAAAAAAGAGGAGGCGGGLGSGTSGRGLGGGAGGAGGAGGAGGAGGAGGAGGAG 450  
DB 371 AG-----AVAAAAAGGAGGCGGGLGSGT--AGRGGAG-----AAAA 419  
QY 451 LGSOTSGRGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 497  
DB 420 LGNGP--AGRGGLGGGAG-----AAAAAGGAGGCGGGLGSGTSGRGLGGGAG 475  
QY 498 GYGIP--GQGTSGRGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 550  
DB 476 GYGGLGSGQAGR--GGCGAGAAAAAVGAGGAGGAGGAGGAGGAGGAGGAGGAG 531

QY 551 GGCGAGAAAA-----AAAAAGGAGGCGGGLGSGTSGRGLGG 590  
DB 532 GGCGAGAAAAAAGGAGGCGGGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 587  
QY 591 QGAGAAAAAAGAGGAGGCGGGLGSGTSGRGLGGGAGGAGGAGGAGGAGGAGGAGGAG 650  
DB 588 QGAG-----AAAAAGGAGGCGGGLGGGAG--GRGGLGGGAG-----AAAA 636  
QY 651 GGLGS 655  
DB 637 GGVGS 641  
RESULT 4  
US-08-317-844B-2  
Sequence 2, Application US/08317844B  
Patent No. 5989894  
GENERAL INFORMATION:  
APPLICANT: Lewis, Randolph V.  
APPLICANT: Xu, Ming  
APPLICANT: Himman, Michael B.  
TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK  
TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL  
TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF  
NUMBER OF SEQUENCES: 62  
CORRESPONDENCE ADDRESS:  
ADDRESS: Birch, Stewart, Kolasch & Birch  
STREET: 301 No. 5989894th Washington Street  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22046  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/317,844B  
FILING DATE: 04-OCT-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 1447-105P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 241-1300  
TELEFAX: (703) 241-2848  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 718 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-317-844B-2

Query Match 61.1%; Score 2135; DB 2; Length 718;  
Best Local Similarity 68.3%; Pred. No. 7.8e-151;  
Matches 495; Conservative 15; Mismatches 39; Indels 176; Gaps 34;

QY 23 QGAGAAAAAAGAGGAGGCGGGLGSGGT--SGRGLGGGAG--AAAAAAGAGGAG 79  
DB 1 QGAG-----AAAAAGGAGGCGGGLGGGAGGAGGAGGAGGAGGAGGAGGAG 56  
QY 80 GYGGLGSGTSGRGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 139  
DB 57 GYGGLGSGT--AGRGGAG-----AAAAAGGAGGCGGGLGSGGT--AGRG 107  
QY 140 GAAAAAAGAGGAGGCGGGLGSGGTSGPGYGPQGTSGRGLGGGAGGAGGAGGAG 199  
DB 108 G-----AAAAAGGAGGCGGGLGSGT--AGRGGAG-----AAAA 142





Query Match	56.5%	Score 1972.5	DB 4	Length 606
Best Local Similarity	64.2%	Pred. No. 6.7e-139		
Matches 462	Conservative 8	Mismatches 35	Indels 215	Gaps
38	GAGCGGCGGLGSGQTSGRGGTGGCGAGAAAAAAGGAGAGCGGCGGLGSGQTSGRGGTGG 97			
1	GAGCGGCGGLGSGQ-AGRGGTGGCGAG---AAAAAAGGAGGQ-----GGTG 42			
98	GCGAGAAAAAAGGAGCGGCGGLGSGQTSGRGGTGGCGAGAAAAAAGGAGAGCGG 157			
43	SGAGCGGAGAAAAAGGAGCGGCGGLGSGQ---AGRGGGAG---AAAAAGGAGCGG 94			
158	YGGLGSGQTSGRGGYGP-GQGTSGRGGTGGCGAGAAA-----A 195			
95	YGGLGSGQ-AGGCGYGGTGGCGAGRGGTGGCGAGAAAAAGGAGCGGCGGAGCGAG 153			
196	AAAAAGGAGCGGCGYGGTGGSGQTSGRGGTGGCGAGAAAAAAGGAGCGGCGGLGSGQT- 254			
154	AAAAAGGAGCGGCGYGGTGGSGQ---AGRGGGAG---AAAAAGGAGCGGCGGLGSGQAG 205			
255	-----SGRGGTGGCGAGAAAAAAGGAGCGGCGGLGSGQTSGRGGTGGCGAG 303			
206	QGGYGGTGGCGAGRGGTGGCGAG---AAAAAGGAGCGQ-----GGLGSGAG 248			
304	GAATAAAAAAGGAGCGGCGYGGTGGSGQTSGRGGTGGCGAGAAAAAAGGAG 363			
249	GCGAGAAAAAGGAGCGGCGYGGTGGSGQ-----AGRGGGAG---AAAA 289			

RESULT 9  
US-09-247-806-8  
; Sequence 8, Application US/09247806  
; Patent No. 6280747  
; GENERAL INFORMATION:  
; APPLICANT: PHILIPPE, Michel  
; APPLICANT: GARSON, Jean-Claude  
; APPLICANT: ARRAUDEAU, Jean-Pierre  
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT  
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN  
; TITLE OF INVENTION: ANALOG  
; FILE REFERENCE: 6388-0365-0  
; CURRENT APPLICATION NUMBER: US/09/247.806  
; CURRENT FILING DATE: 1999-02-11  
; EARLIER APPLICATION NUMBER: FR 98/01614  
; EARLIER FILING DATE: 1998-02-11  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 606







QY 136 GGGAGAAAAAAGAGAGGCGGGLGSGQTSRGGYGPQGTSGRGLGGGAGAAAA 195  
DB 184 YGGRGAGAGAAA--GAGAGYGGGGYGAGAGAGAAAAAGAGSGGAGGYGRGAGAGA 241  
QY 196 AAAAGAGAGGCGGGLGSGQTSRGGGLGGGAGAGAAAAA-----AAAAAG 239  
DB 242 GAAAGAGAGAGSYG-----GQYGAGAGAGAAAAAAXXXXXXXXXXXGAGAGAG 293  
QY 240 GAGGGYG-----GLGSGQTSRGLGGGAGAGAAAAAAGAGAGGCGG 286  
DB 294 YGGGGYGAGAGAGAAAAAAGAGAGAGGYGRGAGAGAGAAAAAGAGAGGCGG 353  
QY 287 ---LGSQTSRGLG---CGGAGAAAAAAGAGAGGCGGGLGSGQTSRGGYGP 339  
DB 354 GAAAGAGAGAGAGGYGRGAGAGAGAGAGAGAGAGAGAGGCGGCGYG-AGAGAGAAA 412  
QY 340 QOTSGRGLGGGAGAGAAAAAAGAGAGGCGGGLGSGQ-----TSRGGGLGG 388  
DB 413 AAGAGAGAGGYGRGAGAGAGAAA--GAGAGYGGGGYGAGAGAGAAAAATGAGGAG 470  
QY 399 QGAGAAAAAAGAGAGGCGGGLGSGQTSRGLGGGAGAGAAAAAAGAGAGGCGG 448  
DB 471 YGRGAGAGAGAGAGAGAGTGG-----AGYGGGGYGAGAGAGAAAAAGAGAGGAGY 521  
QY 449 G-GLGSGQTSRGLGGGAGAGAAAAAAGAGAGGCGGCGG---LGSQTSRGP- 503  
DB 522 GRGAG-----AGAGAAAGAGAGAGAGAGAGAGGCGGCGYGAGAGAGAGAGAGAG 577  
QY 504 QOTSGRGLGGGAGAGAAAAAAGAGAGGCGGGLGSGQ-----QGTSGRGLGG- 554  
DB 578 YSRGRAGAGAGAGAGAGAGAGAGAGGCGGCGYGAGAGAGAGAGAGAGAGAGAGAG 637  
QY 555 AGAAAAAAGAGAGGCGGGLGSGQTSRGGGLGGGAGAGAAAAAAGAGAGGCGG 614  
DB 638 AGAAGAGAGAGAGAGAGY-----GGGCGYGAGAGAGAAAAAGA--GAGRGY- 683  
QY 615 LGSQTSRGLGGGAGAGAAAAAAGAGAGGCGGCGG 651  
DB 684 -RGAGAGYGGGGYGAGAGAGAGAAAA--GAGAGGYG 717

## RESULT 13

US-08-458-298-2  
; Sequence 2, Application US/08458298  
; Patent No. 5756677  
; GENERAL INFORMATION:  
; APPLICANT: Lewis, Randolph V.  
; TITLE OF INVENTION: CDNAS Encoding Minor Ampullate Spider  
; TITLE OF INVENTION: Silk Proteins  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-3487  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/458,298  
; FILING DATE: 02-JUN-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/209,747  
; FILING DATE: 14-MAR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy Jr., Gerald M.

; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 1447-104P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-205-8000  
; TELEFAX: 703-205-8050  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 832 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; FRAGMENT TYPE: Internal  
; ORIGINAL SOURCE:  
; ORGANISM: N. clavipes  
; TISSUE TYPE: minor ampullate gland  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..309  
; US-08-458-298-2

Query Match 40.2%; Score 1405; DB 1; Length 832;  
Best Local Similarity 50.1%; Pred. No. 6.7e-97;  
Matches 349; Conservative 17; Mismatches 207; Indels 124; Gaps 26;  
QY 21 GQGGAGAAAAAAGAGAGGCGGGLGSGQTSRGGGLGGGAGAGAAAAAAGAGAGGCG 80  
DB 79 GYGAG--AGAVAAAGAGAGGYG---RGAGGYGGGGYGAGAGAGAAAAA--CAGAG 129  
QY 81 GYGGLGSGQTSRGLGGGAGAGAAAAAAGAGAGGCGGCGG---GLGSGQTSRGLGG 135  
DB 130 GAGGY-----CRGAGAGAGAGAGAGAGAGAGAGGCGGCGYGAGAGAGAAAAAGAGAGGAG 183  
QY 136 GQAGAAAAAAGAGAGGCGGGLGSGQTSRGGYGPQGTSGRGLGGGAGAGAAAA 195  
DB 184 YGRGAGAGAGAAA--GAGAGYGGGGYGAGAGAGAGAAAAAGAGAGGAGGYGRGAGAGA 241  
QY 196 AAAAGAGAGGCGGGLGSGQTSRGGGLGGGAGAGAAAA-----AAAAAG 239  
DB 242 GAAAGAGAGAGSYG-----GQYGAGAGAGAAAAAAXXXXXXXXXXXGAGAGAG 293  
QY 240 GAGGGYG-----GLGSGQTSRGLGGGAGAGAAAAAAGAGAGGCGGCGG----- 286  
DB 294 YGGGGYGAGAGAGAAAAAAGAGAGAGAGGYGRGAGAGAGAGAGAGAGGCGGCGY 353  
QY 287 ---LGSQTSRGLG---GQAGAAAAAAGAGAGGCGGGLGSGQTSRGGYGP 339  
DB 354 GAAAGAGAGAGAGGYGRGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 412  
QY 340 QOTSGRGLGGGAGAGAAAAAAGAGAGGCGGGLGSGQ-----TSRGGGLGG 388  
DB 413 AAGAGAGAGGYGRGAGAGAGAAA--GAGAGYGGGGYGAGAGAGAAAAATGAGGAG 470  
QY 389 QGAGAAAAAAGAGAGGCGGGLGSGQTSRGLGGGAGAGAAAAAAGAGAGGCGG 448  
DB 471 YGRGAGAGAGAGAGAGAGTGG-----AGYGGGGYGAGAGAGAGAAAAAGAGAGAGY 521  
QY 449 G-GLGSGQTSRGLGGGAGAGAAAAAAGAGAGGCGGCGG---LGSQTSRGP- 503  
DB 522 GRGAG-----AGAGAAAGAGAGAGAGAGAGAGGCGGCGYGAGAGAGAGAGAGAG 577  
QY 504 QOTSGRGLGGGAGAGAAAAAAGAGAGGCGGCGG---GLGS-----QGTSGRGLGG- 554  
DB 578 YSRGRAGAGAGAGAGAGAGAGAGAGGCGGCGYGAGAGAGAGAGAGAGAGAGAGAG 637  
QY 555 AGAAAAAAGAGAGGCGGGLGSGQTSRGGGLGGGAGAGAAAAAAGAGAGGCGG 614  
DB 638 AGAAGAGAGAGAGAGY-----GGGCGYGAGAGAGAAAAAGA--GAGRGY- 683  
QY 615 LGSQTSRGLGGGAGAGAAAAAAGAGAGGCGGCGG 651  
DB 684 -RGAGAGYGGGGYGAGAGAGAGAAAA--GAGAGGYG 717



Db 193 -----GPSGSAANAAAAAGPGYGPQOGPGYG-PGQGGPGYGPQOGPGSPGSA 246  
Qy 235 AAAGGAGGGYGLGSGTSGRGLG-GQAGAAAAAAGAGGAGGGYGLGSGTSG 293  
Db 247 AAAAAAGPGGPGYG-PGQGGPGYGPQOGPGSPGSAANAAAAAGPGYGP-GQOGPG 304  
Qy 294 GRGGLGGAGAAAAAAGGAGGGYGLGSGTSGRGLGPGQGTSGRGLGGGGA 353  
Db 305 GYGP-GQGGPGPGSAANAAAAAGPGYG----PGQGGPGYGPQOGPGYGPQOGPG 359  
Qy 354 GAAAAAAGGAGGGYGLGSGTSGRGLG-GQAGAAAAAAGAGGAGGGYGLG 412  
Db 360 SGPGSAANAAA-GPGQGGPGYG-PGQGGPGYGPQOGPGSPGSAANAAAAAGPGYGP 417  
Qy 413 LGSQTSRGLGGGAGAAAAAAGGAGGGYGLGSGTSGRG-----GLGGGAG 468  
Db 418 -GQGGPGYGP-GQGGPGPGSAANAAAAAAGPGYGP-GQGGPGYGPQOGPGYGP 474  
Qy 469 -----AAAAAAGGAGGGY-----GLGSGTSG-----GPGY 500  
Db 475 QGGPGGSAANAAAGPGQGGPGYGPQOGPGYGPQOGPGSAANAAAAAGPGY 534  
Qy 501 GPQGTSGRGLGGGAGAAAAAAGGAGGGYGLGSGTSGRG---GLGGGAG 556  
Db 535 GPGQGGPGYGPQGGPGSPGSAANAAAAAGPGYGP-GQGGPGYGPQOGPGYGP 593  
Qy 557 -----AAAAAAGGAGGGYGLGSGTSGRGLG-GQAGAAAAAAGAGGAGGG 611  
Db 594 QGGPGGSAANAAAGPGQGGPGYG-PGQGGPGYGPQOGPGSPGSAANAAAAAGPG 652  
Qy 612 YG-GLGSGTSGRGLGGGAGAAAAAAG-----GAGGGYGLGSGTSGPGYGP 666  
Db 653 YGPQGGPGYGPQOGPGSPGSAANAAAAAAGPGYGPQOGPGYG-PGQGGPGYGP 711  
Qy 667 GQG 669  
Db 712 GQG 714

Search completed: December 18, 2002, 16:23:06  
Job time : 18.6792 secs

**This Page Blank (uspto)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:21:10 : Search time 9.40752 Seconds  
(without alignments)  
1213.493 Million cell updates/sec

Title: US-09-490-291-4  
Perfect score: 3493  
Sequence: 1 MRGSHHHHGHGSMASRGGL.....GGYGGQQTSGIRRPAAKLN 681

Scoring table: BLOSUM62  
Gapop 10.0 , Capext 0.5  
Searched: 106657 seqs, 16763532 residues

Total number of hits satisfying chosen parameters: 106657

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Published Applications\_AA:
- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB pep.\*
  - 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB pep.\*
  - 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB pep.\*
  - 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB pep.\*
  - 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB pep.\*
  - 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep.\*
  - 7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB pep.\*
  - 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep.\*
  - 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB pep.\*
  - 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB pep.\*
  - 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB pep.\*
  - 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB pep.\*
  - 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB pep.\*
  - 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2135	61.1	651	10	US-09-861-597-1 Sequence 1, Appl
2	2076	59.4	606	10	US-09-861-597-4 Sequence 4, Appl
3	1972.5	56.5	606	10	US-09-861-597-8 Sequence 8, Appl
4	1961.5	56.2	606	10	US-09-861-597-6 Sequence 6, Appl
5	1331.5	38.1	714	10	US-09-861-597-10 Sequence 10, Appl
6	1097.5	31.4	529	10	US-09-861-597-2 Sequence 2, Appl
7	861.5	24.7	1894	12	US-10-052-586-97 Sequence 97, Appl
8	738	21.1	4679	10	US-09-804-898-2 Sequence 2, Appl
9	686	19.6	1300	12	US-10-052-586-269 Sequence 269, App
10	675.5	19.3	1497	10	US-09-060-854B-2 Sequence 2, Appl
11	672	19.2	3907	9	US-10-029-217A-24 Sequence 24, Appl
12	656.5	18.8	2211	9	US-10-096-961-1 Sequence 37, Appl
13	652	18.7	1400	10	US-09-879-957-37 Sequence 20, Appl
14	651.5	18.7	720	10	US-09-756-071B-20 Sequence 525, App
15	609	17.4	4440	12	US-10-052-586-525 Sequence 35807, A
16	607	17.4	357	10	US-09-864-761-35807 Sequence 451, App
17	605.5	17.3	1743	12	US-10-052-586-451 Sequence 1, Appl
18	582	16.7	1422	10	US-09-735-933-1 Sequence 3, Appl
19	570.5	16.3	1200	10	US-09-826-508-3

ALIGNMENTS

RESULT 1

US-09-861-597-1  
; Sequence 1, Application US/09861597  
; Patent No. US20020064539A1  
; GENERAL INFORMATION:  
; APPLICANT: PHILLIPPE, Michel  
; APPLICANT: GARRON, Jean-Claude  
; APPLICANT: ARAUDEAU, Jean-Pierre  
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT  
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN  
; TITLE OF INVENTION: ANALOG  
; FILE REFERENCE: 6388-0365-0  
; CURRENT APPLICATION NUMBER: US/09/861,597  
; PRIOR FILING DATE: 2001-05-22  
; PRIOR APPLICATION NUMBER: 09/247,806  
; PRIOR FILING DATE: 1999-02-11  
; PRIOR APPLICATION NUMBER: FR 98/01614  
; PRIOR FILING DATE: 1998-02-11  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 651  
; TYPE: PRT  
; ORGANISM: Nephila clavipes  
US-09-861-597-1

Query Match 61.1%; Score 2135; DB 10; Length 651;  
Best Local Similarity 68.3%; Pred. No. 1e-115;  
Matches 495; Conservative 15; Mismatches 39; Indels 176; Gaps 34;

QY 23 QGAGAAAAAAGAGGAGGGYGLGSGQT-SGRGGLGGGAG-AAAAAAAAAANGAGQ 79  
Db 1 QGAG-----AAAAAGAGGAGGGYGLGSGGAGGAGGAGGAGAGAGAGAGQ 56  
QY 80 GYGGLGSGQTSGRGGGLGGGAGAGAAAAAAGAGGAGGGYGLGSGGTSRGGGLGGGGA 139  
Db 57 GYGGLGSGQ-----AGRGGGAG-----AAAAAGAGGAGGGYGLGSGG-AGRGGLGGGGA 107  
QY 140 GAAAAAAGAGGAGGGYGLGSGGTSRGGGPGGQTSRGGGLGGGAGAGAGAGAGAGAG 199  
Db 108 G-----AAAAAGAGGAGGGYGLGNGQ-----AGRGGQ-----AAAA 142

Sequence 47, Appl  
Sequence 581, Appl  
Sequence 56, Appl  
Sequence 36720, A  
Sequence 37, Appl  
Sequence 37, Appl  
Sequence 34, Appl  
Sequence 34, Appl  
Sequence 39, Appl  
Sequence 39, Appl  
Sequence 35, Appl  
Sequence 35, Appl  
Sequence 38, Appl  
Sequence 38, Appl  
Sequence 22, Appl  
Sequence 14, Appl  
Sequence 431, Appl  
Sequence 19, Appl  
Sequence 19, Appl  
Sequence 18, Appl  
Sequence 18, Appl  
Sequence 7, Appl  
Sequence 25, Appl  
Sequence 36, Appl  
Sequence 25, Appl























```
Db 996 CTGGAGGGCCAGGAGGGAGCCCTCATCTATCATGGCCATACCCTCACCTCCAAGAT 1055
Qy 371 -----GYGLGSGTSGRGLGGGAG----- 392
Db 1056 TCTCTCCGGACGTGGCCCAAGCCGTGCGGACCATGCCTTACGCTGTCCCTTACCC 1115
Qy 393 -----AAAAAAGAGAGGCGGYGGLGSGTSGRGLGGGGA----- 429
Db 1116 TGTATCTCTATCCTGTGAGAACCACTCGGGCTGGAGCAGCA--GGCTGCCATGGCCGC 1173
Qy 430 -----GAAAAAAGAGAGGCGGYGGLGSGTSG--RGLGGGAG-----AAAAA 476
Db 1174 CACCTCTGCACCATCTCTGGGGACATGCTGTGTGACACAGGCGCTGGACTCCCCCAATCCC 1233
Qy 477 AAGAGGCG-----YGLGSGTSGPBGY--PGQTSRGLGGGAGAAAAA----- 524
Db 1234 GAGGAGTGGCATCCCGAGAGCAGCTGAAGGGCCGGTCTGTGTGAGGGAAGAAGCTG 1293
Qy 525 -----AAAAGGAGGCGGYGGLGSGTSGRGLGGGAGAAAAAAGAGAGCGGYGLG 578
Db 1294 CCCGCTCTCGGACGAGATG--GCCG--GGCTGTCTCGGATCGGGAGGGGAGGAG 1350
Qy 579 SGTSGRGLGGGAGAAAAAAGAGAGCGGYGGLGSGTSGRGLGGGAGAAAAA 638
Db 1351 --GATGACGAGGAGGAAGAGAGGCTGGAGGCTGCAGCGCAGAGCGGCTGCCCAAGC 1408
Qy 639 AAA-----AGGAGCG-----YGLGSGTSGP 662
Db 1409 AGATCTCCCGGAGCTGTGGCCCTGTGTACTG 1445

RESULT 13
US-09-879-957-37
; Sequence 37, Application US/09879957
; Patent No. US20020034755A1
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; HOFFMAN, No. US20020034755A1h
; KAY, Brian K.
; FOWLKES, Dana M.
; MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/879,957
; FILING DATE: 13-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/630,915
; FILING DATE: 03-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 37:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1400 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 37:
US-09-879-957-37

Query Match 18.7% Score 652; DB 10; Length 1400;
Best Local Similarity 28.3%; Pred. No. 5.2e-31;
Matches 274; Conservative 24; Mismatches 325; Indels 344; Gaps 32;

Qy 14 ASGRGGLGGCGAGAAAAAAGAGAGCGGYGGLGSGTSGRGLGGGAGAAAAA 73
Db 2 AAAGGAGGAGAGTGTCAAAAA--GAGGATGGCGA-----GGAAAAAGGCACACA 50
Qy 74 AGAGGCGGYGGLGSGTSGRGLGG-----OGAGAAAAAAGAGAGCG 118
Db 51 GGAAGCACAGACAGCTGGTGGCTTTTCCATCAACACCAAGAACCCAGCTAAGCCA-- 108
Qy 119 GYGGLGSGTSGRGG-----LGQ-----GAGAAAAA----- 145
Db 109 -----GCTGTCCAGGCACCCCTGCTCCACTGCAGAAAAAGGCTCCACTTTACCA 163
Qy 146 -----AAAAAGAGCGGYGGLGSGTSGPBG-----YGPQQTSGRGG---L 184
Db 164 CAGAAAAATGTAAAGTGGTGTATTACCGGCACTGTACCCCTTTGATCCAGAGCCAT 223
Qy 185 GGCGAGAAAAAAGAGAGCGGYGGLGSGTSGRGLGGGAGAAAAAAGAGAGCG 244
Db 224 GATCAATCACTATCCAGCCAGAGACATAGTCATGTTGGATGAAGCAACACATGGAGAA 283
Qy 245 GYGGLGSGTSGRGLGGGAGAAAAAAGAGAGCGGYGGLGSGT-----SGRG 296
Db 284 CCCG--GCTGGCTTGGAGGAGAAATTAAGGAAAGACAGG----GTGGTTCCCTGCAAACT 338
Qy 297 GLGGGAGAAAAAAGAGAGCGGYGGLGSGTSGPBGYPGQTSRGLGGGAGAA 356
Db 339 ATGCAGAGAAAAATCCAGAAATAGGTTCCCGCTCCAGTGAACACCACTGACTGATTCNA 398
Qy 357 AA-----AAAAAGAGAGCGGY--GLG-----QGT----- 380
Db 399 CATCTGCCCTCCCCCAAACTGGCTTGCTGAGACCCCGCCCTTTGGCAGTAACCT 458
Qy 381 -----SGRG-----GLGG----- 388
Db 459 CTTCAGAGCCCTCCACGACCCTTAATAACTGGCGCGCACTTTCAGCTCCAGCTGGCCACCA 518
Qy 389 -----QGAGAAAAAAGAGAGCGGYGGLGSGTSGRGLGG-----QGAGAAAAA 438
Db 519 GCACGAATGAGAAACCAAGAACGGATAACTGGGATGATGGCAGCCAGCCCTCTCTCA 578
Qy 439 AAG-----GAGCGGYGGLGSGTSGRGLGG-----AGAAAAA----- 477
Db 579 CCGTTCCAAGTGGCCGC--CAGTTAAGCAGAGGTCGCCCTTTACTCCAGCCAGCCCACT 637
Qy 478 -----AGGAGCGGYGGLGSGTSGPBGYPGQTSRGG--- 511
Db 638 GGCTCCTCCCGTCTCTGTGCTAGGCCAG--GGTGAAGAGTGGAGGGCTACAAGCTC 695
Qy 512 -----LGCGAGAAAAA----- 528
Db 696 AAGCCCTATATCTTTGGAGAGCCAAAAAGACAAACCACTTAATTTTAAACAAAATGATG 755
Qy 529 -----GGAGCGGYGGLGSGTSGRGLGGGAGAA---AAAAAAGGAG 571
Db 756 TCATCAGCGTCTCGAAACAGACACATGTTGGTTGGAGAGTTCAAGGTGAGAGG 815
Qy 572 GYGGLG-----SQGT-----SGRGLGGGAGAAAAAAGAG 608
Db 816 GTTGGTTCCCAAGCTCTTACGTGAACACTCAATTCAGGCCCATTAAGGAAGTCTACAAGCA 875
```



1	PRIOR APPLICATION NUMBER: 60/063541	
2	PRIOR FILING DATE: 1997-10-28	
3	PRIOR APPLICATION NUMBER: 60/063544	
4	PRIOR FILING DATE: 1997-10-28	
5	PRIOR APPLICATION NUMBER: 60/063564	
6	PRIOR FILING DATE: 1997-10-28	
7	PRIOR APPLICATION NUMBER: 60/063734	
8	PRIOR FILING DATE: 1997-10-29	
9	PRIOR APPLICATION NUMBER: 60/063870	
10	PRIOR FILING DATE: 1997-10-31	
11	PRIOR APPLICATION NUMBER: 60/064103	
12	PRIOR FILING DATE: 1997-10-31	
13	PRIOR APPLICATION NUMBER: 60/065311	
14	PRIOR FILING DATE: 1997-11-13	
15	PRIOR APPLICATION NUMBER: 60/066120	
16	PRIOR FILING DATE: 1997-11-21	
17	PRIOR APPLICATION NUMBER: 60/066466	
18	PRIOR FILING DATE: 1997-11-24	
19	PRIOR APPLICATION NUMBER: 60/066772	
20	PRIOR FILING DATE: 1997-11-24	
21	PRIOR APPLICATION NUMBER: 60/069335	
22	PRIOR FILING DATE: 1997-12-11	
23	PRIOR APPLICATION NUMBER: 60/069425	
24	PRIOR FILING DATE: 1997-12-12	
25	PRIOR APPLICATION NUMBER: 60/069870	
26	PRIOR FILING DATE: 1997-12-17	
27	PRIOR APPLICATION NUMBER: 60/068017	
28	PRIOR FILING DATE: 1997-12-18	
29	PRIOR APPLICATION NUMBER: 60/077450	
30	PRIOR FILING DATE: 1998-03-10	
31	PRIOR APPLICATION NUMBER: 60/077632	
32	PRIOR FILING DATE: 1998-03-11	
33	PRIOR APPLICATION NUMBER: 60/077649	
34	PRIOR FILING DATE: 1998-03-11	
35	PRIOR APPLICATION NUMBER: 60/078886	
36	PRIOR FILING DATE: 1998-03-20	
37	PRIOR APPLICATION NUMBER: 60/078939	
38	PRIOR FILING DATE: 1998-03-20	
39	PRIOR APPLICATION NUMBER: 60/079664	
40	PRIOR FILING DATE: 1998-03-27	
41	PRIOR APPLICATION NUMBER: 60/079786	
42	PRIOR FILING DATE: 1998-03-27	
43	PRIOR APPLICATION NUMBER: 60/080107	
44	PRIOR FILING DATE: 1998-03-31	
45	PRIOR APPLICATION NUMBER: 60/080194	
46	PRIOR FILING DATE: 1998-03-31	
47	PRIOR APPLICATION NUMBER: 60/080327	
48	PRIOR FILING DATE: 1998-04-01	
49	PRIOR APPLICATION NUMBER: 60/080333	
50	PRIOR FILING DATE: 1998-04-01	
51	PRIOR APPLICATION NUMBER: 60/081049	
52	PRIOR FILING DATE: 1998-04-08	
53	PRIOR APPLICATION NUMBER: 60/081070	
54	PRIOR FILING DATE: 1998-04-08	
55	PRIOR APPLICATION NUMBER: 60/081195	
56	PRIOR FILING DATE: 1998-04-09	
57	PRIOR APPLICATION NUMBER: 60/081838	
58	PRIOR FILING DATE: 1998-04-15	
59	PRIOR APPLICATION NUMBER: 60/082568	
60	PRIOR FILING DATE: 1998-04-21	
61	PRIOR APPLICATION NUMBER: 60/082569	
62	PRIOR FILING DATE: 1998-04-21	
63	PRIOR APPLICATION NUMBER: 60/082704	
64	PRIOR FILING DATE: 1998-04-22	
65	PRIOR APPLICATION NUMBER: 60/082797	
66	PRIOR FILING DATE: 1998-04-22	
67	PRIOR APPLICATION NUMBER: 60/083322	
68	PRIOR FILING DATE: 1998-04-28	
69	PRIOR APPLICATION NUMBER: 60/083495	
70	PRIOR FILING DATE: 1998-04-29	
71	PRIOR APPLICATION NUMBER: 60/083496	
72	PRIOR FILING DATE: 1998-04-29	
73	PRIOR APPLICATION NUMBER: 60/083499	
74	PRIOR FILING DATE: 1998-04-29	

;	PRIOR FILING DATE:	1998-04-29	
;	PRIOR APPLICATION NUMBER:	60/083559	
;	PRIOR FILING DATE:	1998-04-29	
;	PRIOR APPLICATION NUMBER:	60/084366	
;	PRIOR FILING DATE:	1998-05-05	
;	PRIOR APPLICATION NUMBER:	60/084414	
;	PRIOR FILING DATE:	1998-05-06	
;	PRIOR APPLICATION NUMBER:	60/084639	
;	PRIOR FILING DATE:	1998-05-07	
;	PRIOR APPLICATION NUMBER:	60/084640	
;	PRIOR FILING DATE:	1998-05-07	
;	PRIOR APPLICATION NUMBER:	60/084643	
;	PRIOR FILING DATE:	1998-05-07	
;	PRIOR APPLICATION NUMBER:	60/085573	
;	PRIOR FILING DATE:	1998-05-15	
;	PRIOR APPLICATION NUMBER:	60/085579	
;	PRIOR FILING DATE:	1998-05-15	
;	PRIOR APPLICATION NUMBER:	60/085580	
;	PRIOR FILING DATE:	1998-05-15	
;	PRIOR APPLICATION NUMBER:	60/085582	
;	PRIOR FILING DATE:	1998-05-15	
;	PRIOR APPLICATION NUMBER:	60/085700	
;	PRIOR FILING DATE:	1998-05-15	
;	PRIOR APPLICATION NUMBER:	60/086023	
;	PRIOR FILING DATE:	1998-05-18	
;	PRIOR APPLICATION NUMBER:	60/086392	
;	PRIOR FILING DATE:	1998-05-22	
;	PRIOR APPLICATION NUMBER:	60/086486	
;	PRIOR FILING DATE:	1998-05-22	
;	PRIOR APPLICATION NUMBER:	60/087098	
;	PRIOR FILING DATE:	1998-05-28	
;	PRIOR APPLICATION NUMBER:	60/087208	
;	PRIOR FILING DATE:	1998-05-28	
;	PRIOR APPLICATION NUMBER:	60/087609	
;	PRIOR FILING DATE:	1998-06-02	
;	PRIOR APPLICATION NUMBER:	60/087759	
;	PRIOR FILING DATE:	1998-06-02	
;	PRIOR APPLICATION NUMBER:	60/087827	
;	PRIOR FILING DATE:	1998-06-03	
;	PRIOR APPLICATION NUMBER:	60/088025	
;	PRIOR FILING DATE:	1998-06-04	
;	PRIOR APPLICATION NUMBER:	60/088028	
;	PRIOR FILING DATE:	1998-06-04	
;	PRIOR APPLICATION NUMBER:	60/088029	
;	PRIOR FILING DATE:	1998-06-04	
;	PRIOR APPLICATION NUMBER:	60/088033	
;	PRIOR FILING DATE:	1998-06-04	
;	PRIOR APPLICATION NUMBER:	60/088167	
;	PRIOR FILING DATE:	1998-06-05	
;	PRIOR APPLICATION NUMBER:	60/088202	
;	PRIOR FILING DATE:	1998-06-05	
;	PRIOR APPLICATION NUMBER:	60/088212	
;	PRIOR FILING DATE:	1998-06-05	
;	PRIOR APPLICATION NUMBER:	60/088217	
;	PRIOR FILING DATE:	1998-06-05	
;	PRIOR APPLICATION NUMBER:	60/088326	
;	PRIOR FILING DATE:	1998-06-04	
;	PRIOR APPLICATION NUMBER:	60/088655	
;	PRIOR FILING DATE:	1998-06-09	
;	PRIOR APPLICATION NUMBER:	60/088722	
;	PRIOR FILING DATE:	1998-06-10	
;	PRIOR APPLICATION NUMBER:	60/088738	
;	PRIOR FILING DATE:	1998-06-10	
;	PRIOR APPLICATION NUMBER:	60/088740	
;	PRIOR FILING DATE:	1998-06-10	
;	PRIOR APPLICATION NUMBER:	60/088811	
;	PRIOR FILING DATE:	1998-06-10	
;	PRIOR APPLICATION NUMBER:	60/088824	
;	PRIOR FILING DATE:	1998-06-10	
;	PRIOR APPLICATION NUMBER:	60/088825	
;	PRIOR FILING DATE:	1998-06-10	
;	PRIOR APPLICATION NUMBER:	60/088826	
;	PRIOR FILING DATE:	1998-06-10	

:	PRIOR APPLICATION NUMBER: 60/088861	
:	PRIOR FILING DATE: 1998-06-11	
:	PRIOR APPLICATION NUMBER: 60/088863	
:	PRIOR FILING DATE: 1998-06-11	
:	PRIOR APPLICATION NUMBER: 60/088876	
:	PRIOR FILING DATE: 1998-06-11	
:	PRIOR APPLICATION NUMBER: 60/089090	
:	PRIOR FILING DATE: 1998-06-12	
:	PRIOR APPLICATION NUMBER: 60/089105	
:	PRIOR FILING DATE: 1998-06-12	
:	PRIOR APPLICATION NUMBER: 60/089512	
:	PRIOR FILING DATE: 1998-06-16	
:	PRIOR APPLICATION NUMBER: 60/089514	
:	PRIOR FILING DATE: 1998-06-16	
:	PRIOR APPLICATION NUMBER: 60/089538	
:	PRIOR FILING DATE: 1998-06-17	
:	PRIOR APPLICATION NUMBER: 60/089598	
:	PRIOR FILING DATE: 1998-06-17	
:	PRIOR APPLICATION NUMBER: 60/089653	
:	PRIOR FILING DATE: 1998-06-17	
:	PRIOR APPLICATION NUMBER: 60/089908	
Query Match		17.4%; Score 609; DB 12; Length 4440;
Best Local Similarity		26.0%; Pred. No. 3.3e-28;
Matches 283; Conservative		27; Mismatches 332; Indels 446; Gaps
Qy	15	SGRGLGG---OCAGAAAAAAGACAGQG---GVGGLGSQGT--SGRGLGGGCGAGAA 66
Db	152	TGTCTGGCATTTGTGTATCGATCGCATGTGTGCGTGCACACGTGTGTGTTTGCCTGTGA 211
Qy	67	AAAAAAGGAGGGYGGGLGSQGTSGRGLGGGAGAAAAAAGGAGGQGGYGGGLSQ 126
Db	212	CATCTGCATGTGT---GTGTTGT-GTGCCTGTGTGTGCACATGTAGAAA-GAAAGTGAT 265
Qy	127	GTSGRGLGGGAGAAAAAAGGAGGQGGYGGGLSQG---TSPGGYGPQQTSGRGL 184
Db	266	GTGTGGGAATGGAGATTCCAAACCAGGAGGAGACTGTGCCTGGGCTGCCACGAGAAT 325
Qy	185	GGQCA-----GAAAAA-----AAAAAGGAGGGYG-----GLGSQGT 216
Db	326	GGTGACACTTCGAGAAAGGGTTCCAAATGCACGTTCCTCCAGTTCCTGTGTGACCT 385
Qy	217	SGRGLGG-----GAGAG-----GAG-----AAAA 233
Db	386	TGGCAAGTCACTTCAGCTCCCTGAGCTTTGTTTTAAAAATATTTTTAAATGTATAA 445
Qy	234	AAAAAGGAGGGYGGGLGSQGTSGRGLG---GGAGAAAAAAGAG-----GAGQ 281
Db	446	ACCATGGACCATTACATATGAAGAGAAATGTGTGTCAACATTCATTAANTACAA 505
Qy	282	GGYGGGLGSQGT-----SGRGLSG---QGAGAAAAAAGAGAG---Q 319
Db	506	GGTGGAGAGTGCTGTCTACAGCCAGAGCTGACACAGGGACTTCGCCAGAGGCTGCA 565
Qy	320	GG-----YGLGSQGTSGPGYG-----PGQTSRGR 346
Db	566	GGGCTTCCCTCCCAACCTTCCACGGGCCCACTTGCTGAGCCCGTGCTATCTGTTCAGTG 625
Qy	347	GLGSGQ-----AGAAAA-----AAAGGAGQG-GYG-----GLG 376
Db	626	GAAATGCCATGTGCCAGTTCCTCAGAACTTAAGCCACTAGTAAAGTGTGACACTCTGTG 685
Qy	377	-----SQGTSGRGLG-----GGAG-----AAAAAAGAGCA 405
Db	686	CCAGCTTGTGTCAAGCCCTTGTGTGTGCTATCTATTATGACCTCCCAATAGCATGTG 745
Qy	406	GQGG-YGGLGSQGT-----SGRGLGGGAGAAAA-----AAAAAG 442
Db	746	AAGGTCACTACTGTCACTTCCCTCCCTTCAAGATGAGGAGACCAAAATCTAGATGA 805
Qy	443	AGGGYGGGLGSQGTSGRGG-----LGGGAGAAAAAAGAG----- 478
Db	806	AATGGAGTCCAGCTGGAAATAGGATCCAGACAGATTAAATCCAATTCACAGCCTGAAT 865





Db 352 CKGGVGGVAGLGGAGGAAGQLFSAGGAAGAVGVGCTGCGGCGGAGGAGAGADAPASTGLTG 411  
Qy 90 -----TSRGRGLGGCGAGAAAAAAGCAGCGGCGGGLGSGTSGR-----GGGCG 136  
Db 412 CTGFAGGAGGCGGGGNAIAGGINSGGAGGTGCGGCGAGGCGADNAGSIGADGGAGG 471  
Qy 137 QG--AGAAAAAAGG-----AGCGGGLGSGGT----- 166  
Db 472 TGGNAGAGGAGCACTGTGTGTVGAAGKAGTGGTGGCGGAGGAGSAGTATATGATGTG 531  
Qy 167 -----SCPGGCPGQQTSGRGLGGCGAGAAAAAAGAGAGQ-----GCGGGLGSGQ 215  
Db 532 FSGGAGGAGGAGGNTGVTGTTNGSGQG-GTGAGGAGGAGGAGGADNPTGIGTGTGCKG 590  
Qy 216 TSRGRGLGGQ--AGAAAAAAGG--ACGGGCGGLGSGT---SRRGLGGCGAGAAA 269  
Db 591 --GAGGAGCGGSSGAGGTNGSGGAGTGGCGGAGGAGGADNPTGCGGAGGTGCTGCA 648  
Qy 270 AAAAAAGGAGGCGG---LGSQTSRGLGGQ-CAGAAAAAAGGA--- 317  
Db 649 ACAGGAGCAITGTGGAVGVNAGTGTGTGGVAGGAGAAAAAGSSATGAGGAG 708  
Qy 318 QGGGCGGLGSGQ---GTSPPGCGPQQTSGRGLGGQCA-----GAAAAA 360  
Db 709 GAGGEGAGGNSGVTGTTNGSGAGGAGGCGTGGAGGSGADNPTCAGFAGGAGGTGGAAG 768  
Qy 361 AAAAGGA--GCGGCGGL-----GSGTSGRGLGGQCA----- 391  
Db 769 AGGAGGATGTGTGTVGATGATGAGTGGAGRGDGDGASGLGLSLGSGDGGCGGCGAG 828  
Qy 392 GAAAAAAGGAGGA--GCGGGLGSGTSG--RGLGGQ--GAGAAA----- 433  
Db 829 CSAGAGGTNGAGGAGGCGGCGATGAAAGLNDGNGVGGGCGGAGGAGGAGNAGVGLTAK 888  
Qy 434 AAAAAAGGAGGCGGGLGSGQ---TSRGLGGQCA---GAAAAAAGGA--- 481  
Db 889 AGDGAAGNGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 948  
Qy 482 CCGGCGGLGSGTSGPCCGCGPQQTSGRGLGGCGAGAAAAAAGGAGGCGGGLGSG 541  
Db 949 CTGCKGAGGAGTGLGVGCGS---GTGGDGDAGSGGGGFGGAGAGGAGGAGGAGGAGG 1005  
Qy 542 QGTS-----GRGLGGCGAGAAAAAAGGAGGCGGGLGSGTSG---RGG 587  
Db 1006 DCAAGLGLGSLGFGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1065  
Qy 588 LGGQ--GAGAAA-----AAAAAAGGAGGCGGGLGSGQ---TSRGLGGQ 629  
Db 1066 VGGDGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1125  
Qy 630 GA---GAAAAAAGGA---GCGGCGGLGSGTSGPCCGCG 665  
Db 1126 GCGGGLGAGTSTINANGAGGNGGTGCKGAGGAGTGLGVGSG 1169

## RESULT 4

D70807

hypothetical glycine-rich protein Rv3514 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C:Accession: D70807

R:Cole, S.T.; Broech, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: D70807

A&gt;Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1489 &lt;COL&gt;

A:Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAA17751.1; PID:g292445

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: Rv3514

C:Superfamily: collagen alpha 1(IV) chain

Query Match 35.9%; Score 1254.5; DB 2; Length 1489;  
Best Local Similarity 39.5%; Pred. No. 2.2e-63;  
Matches 336; Conservative 36; Mismatches 267; Indels 211; Gaps 36;

Qy 16 GRGGLGGQ-----AGAAAAAAGAGAGQ-----GGYGLGSGQTSRGLGGCGAGCA 65

Db 260 GHGCVGGDWLAPGAGGAGGCGGAGGAGSDGALGCTGTG--GTGAGGAGGRGALL 317

Qy 66 AAAAAAAGGA-GGGGCGGLGSGQ-----GTSRGLGG-----QGAGAAAAAAG 113

Db 318 LGAGGCGGLGAGGCGCTCGAGGCGVLGCGVGTGCGGCGVAGLGGAGGAAGLFSASG 377

Qy 114 GAGCGGCGGLGSGQTSRGLGGQCA-----GAAAAAAGGA-----GCGG 157

Db 378 AAGNAGVAGAGGQ--GDGAGGAGADADQPCNTGCTGFAGGAGGAGGAGSSGAGGTNG 435

Qy 158 YGGLGSGQTSRPGGYPGQQTSGRGLGGCGAGAAAAAAGAGCGGCGYGG-LGSGQCT 216

Db 436 SGGAGCGGAGGAGGAGADNPTGIGTGGDGTGAGAGGAGGAAGTGGTGMGTG 495

Qy 217 SRRGLGGQ-----GAGAAA-----AAAAAGGA-----GCGGCG 248

Db 496 AGVCGAGCGGCGGAGGAGADADQPCATCGCTGFAGGAGGAGGAGGAGTNGSGGAGG 555

Qy 249 LGSQ-----TSRGLGGCGAGAAAAAAGAGAGCGGCGYGG-LGSGQTSR 295

Db 556 TGGCGGAGGAGGAGADNPTGIGTGGDGTGAGAGGAGGAAGTGGTGMGTG 615

Qy 296 GGLGSG-----GAGAAAAA-----AAAGGAGQ-----GCGGCG 327

Db 616 GGAGCGGCGGAGGAGADADQPCATCGCTGFAGGAGGAGGAGSSAGTNSGAGGTGR 675

Qy 328 Q-GTSPPGCGPQQTSGRGLGGCGAGAAAAAAGAGCGGCGYGG-LGSGQTSRGG 385

Db 676 QSGTGGAGGAGADNPT-GIGCTGCGDGTGGAAGAGGAGCAAGTGGTGMGTG 734

Qy 386 LGGCGAGAAAAAAGAGGCGGCGGLGSGQ-----SRRGLGGCGAGAAAAAAGG 442

Db 735 AGSSGAGGTNGSGAGGTGCGGAGGAGGAGADNPTGLGTTGGDGTGGAAGAGGAGGA 794

Qy 443 AQCGYGG-LGSQTSRGLGGQ-----GAGAAAAA-----AAAAAGGA----- 481

Db 795 AGTGTGCMGTGTGAGYGGAGGCGGCGGAGGAGADADQPCATGCTGFAGGAGGAGSGG 854

Qy 482 -----GCGGCGGLGSGQTSRGGPCCG-YGCGQQ 505

Db 855 SSCAGGTNGSGAGGTCCQVAGGAGISFNSGSGTGTGCGVG--GTGGDGNAGTCAG 912

Qy 506 TSGRGLGGCGAGAAAAAAGAGG--QGGYGGGLGSGTSGRGLGGCGAGAAAAAAG 564

Db 913 DPKGCTGCTG--GTGCGGAGGSGGNFNGTGGTG--GTGKGLNTDGLSSNTSGTGG 969

Qy 565 AAGGAGCGGCGGLGSGQTSRGLGGCGAGAAAAAAGGA-----GCGGCGGLGSGQ 618

Db 970 TGGTGGKGTGGAGDSDSAGGTGTTGAGGAGGAGGLANTGTTAGNAGTGGDGGCGGNGQ 1029

Qy 619 GTSRGLGGQ-----GAGAAAAA-----AAAAAGAGCGGCGGGLG-----S 655

Db 1030 GDSG--SGLGGPGFAGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1088

Qy 656 GGTSGPGCG 665

Db 1089 GGTGTGGVG 1098

RESULT 5

B70807

hypothetical glycine-rich protein Rv3512 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: B70807  
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome A:Reference number: A70500; MUID:98295987; PMID:9634230  
A: Accession: B70807  
A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
A: Molecule type: DNA  
A: Residues: 1-1079 <COL>  
A: Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAAL17749.1; PID:g292444  
A: Experimental source: strain H37Rv  
C: Genetics:  
A: Gene: Rv3512  
C: Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
Query Match 35.2%; Score 1228.5; DB 2; Length 1079;  
Best Local Similarity 40.0%; Pred. No. 5e-62;  
Matches 318; Conservative 25; Mismatches 303; Indels 149; Gaps 29;  
QY 14 ASGRGGLGGGACAAAAAAGGACGCGYGLGSGQTSRGLGGGAGAAAAA 73  
DB 38 AGGAGAGGTG-GTGGAAGTGTGGGQNGGNG--GTGGKGTGTDGALAGSSGGAGG 94  
QY 74 AGG-----AGCGYGGG-----AGCGYGGG-----SQTSGRGLGGQ 99  
DB 95 KCGNGDGACAGTGSAPGTAGTGGGKCGNGGIGAGTGPVGTGASGTGSGGAGGT 154  
QY 100 GA-GAAAAAAGGA-----GCGYGGG-----SQTSGRGLGGQ----- 137  
DB 155 GGDGAANGGTAGAGGNGGKGGGAGTSTAGNSGAGGSGKGDAGAGGAGAT 214  
QY 138 -----GAGAAAAAAGGACGCGYGLGSG-----GTSGPGYGPQQTSG 180  
DB 215 PGANGTAGNGDGDGAAGAVTSGATGAGDGGHGTGAAGNGGTGGAGGIDYGGG 274  
QY 181 RGLGGLGGGAGAAAAAAGAGAG-OGYGLGSG-----GTSGRGLGGQ 225  
DB 275 TGTGTGNGNGALCGGAGDAGSGNSGNGIGKGNAGAGGANGTGVANGTGGD 334  
QY 226 G--AGAAAAAAGAGAGGCGYGLGSGQTSRGLGGQ-----AGAAAAA 272  
DB 335 GNGGAAGAATAGSNGAGTGSAGNG--GTGGRGSGGAGGDIYGVGKGNADGE 392  
QY 273 AAAGGAG-----OGYGLGSGQTSRGLGGGAGAAAAAAGGACGCGYGLG 325  
DB 393 VGAGGAGGSGPNTSPGNGGQGGGSGGAG--GAAGAGGAGGANGTAGNGGQAGGT 451  
QY 326 GSQGTSGP---GGYGPQQTSGRGLGGGAGAAAAAAGGAGGCGYGLGSGQTS 382  
DB 452 GGAGAAASSATNGSGGAGGTGGDGGSGAGTGCAGGTGGAADGGGCGGAGG--GAGG 510  
QY 383 RGLGGQ-----AGAAAAAAGG--AGCGYGLGSG-----GTSGRGLGGQGA 429  
DB 511 QCGAGGAGGTGGNGNITGTTAGTAGAAGNGGAAGKGGAGGCGGTGGGTGGCGGAGG 570  
QY 430 GAAAAAAGAGAGCGYGLGSGQTSRGLGGGAGAAAAAAGGACGCGYGLG 489  
DB 571 AGTGTGDRTVGGTVPAGSGGQ--GNAGGGGAGG--GGADGGSGDGDAGTGGNGN 626  
QY 490 G-----SQTSGPGYGPQQTSGRGLGGGAGAAAAAAGGAGGCGYGLGSG 542  
DB 627 GGNRNSGNTGAGGNGGCGAGAGGSGGTGNGGAGCGDAGDAGNGNGNTGNG 686  
QY 543 GTSRGLGGGAGAAAAAAGGAGCGYGL-----GSQGT--SGRGLGGGAGAAA 597  
DB 687 GNGNGGIAGMGNCGAGTGSNGNGGSGGNGGNGAGMGNSTGSGDGGAGNGGAAGT 746  
QY 598 AAAAAAGGAGCGYGLGSGQTSRGLGGGAGAAAAAAGGACGCGYGLGSGQ 657

DB 747 GGTGGDGLTGTGTGGSG--GTGGDGGNGGADNTANMTAQAGDGGNGGCGFGGA 804  
QY 658 TSGPGGVGPQQTSG 672  
DB 805 GAGGGLTAGATG 819  
RESULT 6  
A70934  
hypothetical glycine-rich protein Rv0578c - Mycobacterium tuberculosis (strain H37Rv)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: A70934  
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome A:Reference number: A70500; MUID:98295987; PMID:9634230  
A: Accession: A70934  
A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
A: Molecule type: DNA  
A: Residues: 1-1306 <COL>  
A: Cross-references: GB:AL123456; NID:g3242298; PIDN:CAAL17449.1; PID:g290  
A: Experimental source: strain H37Rv  
C: Genetics:  
A: Gene: Rv0578c  
C: Superfamily: collagen alpha 1(IV) chain  
Query Match 34.3%; Score 1197; DB 2; Length 1306;  
Best Local Similarity 38.1%; Pred. No. 3.2e-60;  
Matches 325; Conservative 38; Mismatches 296; Indels 194; Gaps 35;  
QY 3 GSHHHHHG-----SWASRGGLGGGAGAAAAAAGG-----AGCGYGLGSGQTS 53  
DB 388 GSAHGANGAPGTTSTSGNGDGGKADAISSGTCANGRGDGGQVNGGAG--GAG 445  
QY 54 GRGLGGGAGAAAAAAGG--AGCGYGG--LGSQTSRGLGGGCA----- 101  
DB 446 GRGAGLGFSGAPRPGAGGTGGAGNGGTQAGDGTGAGAGGDDGGSGAGSIGF 505  
QY 102 GAAAAAAGG-----AGCGYGL-----GSQTSRGLGGQGA----- 139  
DB 506 NASAPCAAGSPGNGNGPGAGGEGAGGAGLALASGQNGSQAGSDGAGCGTGGN 565  
QY 140 ---CAAAA-----AAAAAGGACGCGYGLGSGQTSG--PGGYG-PGQ--OTSRRGLGG 186  
DB 566 GCHCAAGALGVNGVGGAGGHDGDPGVGGAGCGGSGSTPGANGAPCNTPTSGNGNGG 625  
QY 187 QGAGAAAAAAGG-----AGCGYGLGSG-----GTSGRGG 221  
DB 626 RGADATGCGTGAAGRGDGLVNGGAGGAGCGNGSKGLPLGLRNLNGLDGTGTGNGG 685  
QY 222 LCGGAGAAAAAAGG--AGCGYGLGSG-----GTSGRGLGGQ-- 264  
DB 686 AGSGGAWAGNGTGGAGGTGGYGGTGGVSSAGADGHPGTGGVGTGGKGGD 745  
QY 265 AGAAAAAAGGACGCGYGLGSGQTSRGLGGGAGAAAAAAG--GAGCGGYG 323  
DB 746 GCDGGAAPNGVAGSQPGGAGCDG--GTGGVNGNGRGIDGADGATAGARGDGGAGG 803  
QY 324 GLGSG--GTSPPGYGP-----GQTSRGLGGGQ----- 352  
DB 804 GKGRGTGPGGAGPAGTTGSGCAGNGSGSGTGDGPDGNGANGSVFTNNGICNGG 863  
QY 353 -AGAAAAAAGGA-----GCGYGLGSGQTSRGLGGCGAGAA 394  
DB 864 NGCNAGPSGAGSGGAGTGTGATGSSSSIHVNGNGNGGNDHALSGNCAAGNGNGG 923  
QY 395 AAAAAAGGACGCGYGLGSGQTSRGLGGGAGAAAAAAGGACGCGYGG----- 450





QY 124 -----GSGTSGRGLGQGA-----GAAAAAA 147  
Db 630 AFKFDGGAGGDDGPGAGGKAGGAGATGVTGATGATVHSGNGKRGKGCADATVAG 689  
QY 148 AAAGAGGQGGGL-GSQGTSGPGYG-----PG---QQTSGRG 182  
Db 690 ANGKGGAGGGLVGDGAGGDDGGGAAAGANGNVGDDGDTLSGQPGGSEANGGQ 749  
QY 183 GLGGGAGAAAAAAGAGAGGQGGYGGGLGSGQTSRGLGGGCGAAAAAAGAGAG 242  
Db 750 GYGGGAGGAGGAGGAGSALSGGNGRGDAGQAG----GAGGAGGAGGAGSVSDGG 805  
QY 243 QGGYGLGSGQTSRGLGGGCGAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 301  
Db 806 PGKGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 862  
QY 302 G--AGAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 347  
Db 863 GAGGAGNAPAGQVSHGVGGVGGDGLGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 920  
QY 348 LGG-CCAGAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 385  
Db 921 LGCDGSGNTRAASGVADSDHDPGSGGNGGNGGAGGAGGAGGAGGAGGAGGAGGAGGAG 980  
QY 386 LGGGAGAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 435  
Db 981 AGGNGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1040  
QY 436 AAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 486  
Db 1041 DGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1100  
QY 487 GG-----LGSQGTSGPGYG-----PQQTS 507  
Db 1101 GCAAPSGTVGSHGTSGVGGDGLGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1158  
QY 508 GRGGLGGQ-----GAGAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 543  
Db 1159 GRGGLGGGSGNCTSAANGYDASKHPLTGGDGGVGGNGAKAAAAAGDGGQGGD--GNAG 1216  
QY 544 TSGRGLGGGAGAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 593  
Db 1217 LFGDGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1276  
QY 594 -----GAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 644  
Db 1277 VAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1336  
QY 645 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 672  
Db 1337 NGGGGTGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1364

RESULT 9  
H70846  
hypothetical glycine-rich protein Rv3345c - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: H70846  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: H70846  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1538 <COL>  
A:Cross-references: GB:AL021841; GB:AL123456; NID:g3261517; PIDN:CAA17117.1; PID:g326151  
A:Experimental source: strain H37RV  
C:Genetics:

A:Gene: Rv3345c  
C:Superfamily: collagen alpha 1(IV) chain  
Query Match 33.1%; Score 1157; DB 2; Length 1538;  
Best Local Similarity 34.4%; Pred. No. 6.2e-58;  
Matches 317; Conservative 36; Mismatches 308; Indels 260; Gaps 31;  
QY 1 MRGSHHHHHHSGMASRGGGLGGGAGAAAAAAGG-----AGOGGYGGLGSGQTS 53  
Db 549 LNSAGANGTAPTISGNGGNGGAGATPTVAGENGAGGNGHGGSVGNGAGGAGGNGVA 608  
QY 54 G-----RGLGGGGA-GAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 91  
Db 609 GTGLALNGGNGGNGGIGGNGGSAAGTGGGKGGNGAGANGDFASANGANGGGGNG 668  
QY 92 GRGGLGGGAGAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 139  
Db 669 GNGGIGGKGDAFATFAKAGNGGAGNG-----GNVAVAGGGAGGAGKGAIPANKGATGADG 724  
QY 140 -----GAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 178  
Db 725 TAPTSGDGGNGNGASPTVAGNGDGGKGGSGGNGVNGGNGGAGGAGGAGGAGGAGGAGGAG 784  
QY 179 SRRGLGGGAGAAAAAAGG--AGOGGYGGLGSGQTSRGLGGGAGGAGGAGGAGGAGGAGGAG 236  
Db 785 SGRSSTSGTDGAGGNGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 842  
QY 237 AAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 286  
Db 843 ANGANGENG-----GSGGNGDGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 898  
QY 287 -----LGSQGTSGRGLGQ----- 301  
Db 899 DGGAGENGLNSCAMLPGGCTVGNPCTGNGGNGAGVCGCTGCKAGTGLTGLDGTG 958  
QY 302 -----GAGAAAAAAGG-----AGOGGYG-----LGSQTS 331  
Db 959 PNGNGGNGGNGGKGTAGNGSGAAGGNGGSLNGDAGNGGNGGAGGAGGAGGAGGAGGAGGAG 1018  
QY 332 GPGGY----GPGQQTSGRGLGGGAGAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 387  
Db 1019 GKGNGGNGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1078  
QY 388 GOGAGAAAAAAGGA-----GOGGYGGLGSGQ-----TSRGLGGGAGGAGGAGGAGGAGGAGGAG 436  
Db 1079 GPGAGDFASGVGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1138  
QY 437 AAAAG-----AGOGGYGGLGSGQTSRGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 484  
Db 1139 NGGLGNGGVSETGFGAGGNGGYGPG--GPEGNGGLGNGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1196  
QY 485 GYGLGSGQTSRGGVGPQQTSGRGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 532  
Db 1197 GKGNGGDD--GNVGLGGDAGSGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1254  
QY 533 OGGYGLGSGQTSRGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 588  
Db 1255 SGGAGGNGGTGLNAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1314  
QY 589 GGGGAGAAAAA-----AAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 635  
Db 1315 GGGSSGAAAGSVVNVYTAGHGGNGGNGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1374  
QY 636 -----AAAAAAG-----AGOGGYGGLGSGQTS-----G 660  
Db 1375 DGGNGGNGGNSNSTGVAGLAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1434  
QY 661 PGYGGP-----QQTSGIR 675  
Db 1435 NGGAGAGGSLSTGSGGPR 1455  
RESULT 10

E70917

hypothetical glycine-rich protein Rv1450c - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: E70917  
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome A: Reference number: A70500; MUID: 98295987; PMID: 9634230  
A: Accession: E70917  
A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
A: Molecule type: DNA  
A: Residues: 1-1329 <COL>  
A: Cross-references: GB:295844; GB:AL123456; NID: g3250713; PIDN: CAB09271.1; PID: g2131046  
A: Experimental source: strain H37RV  
C: Genetics:  
A: Gene: Rv1450c  
C: Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology

Query Match 32.8%; Score 1145; DB 2; Length 1329;  
Best Local Similarity 35.9%; Pred. No. 2.6e-57;  
Matches 323; Conservative 35; Mismatches 300; Indels 242; Gaps 35;

```
QY 11 GSMASGRGLGGGCA-----GAAAAAAAAGGA-----GQGGYG 45
DB 139 GGIILWNGGAGGAPQGVGGAGAAGLFTCTGAGGAGGAGGAGGGGWLNGSVG 198
QY 46 GLGSGQTSGRGLGGGAGAAAAAAGAGGOGGYYGLGSGQTSGRGLGGGAGAAA 105
DB 199 GAGQSLIG-GATGAGAGNAGLFGVGTGPGGPGGPGVG--GTGAGGGLGLYGAG 255
QY 106 AAAAAAAGG-AGQGGYGGGLGSQ----GTSRGLGGGAGAAAAA-----AAAA 150
DB 256 HGGAGGPGPIGVGGHGGVGAAGLLGVGGHGGAGHGAEGVAGAACGEDLSPHGTSGVG 315
QY 151 GGAGQGGYGGGLGS--QGTSPGGYGPQQTSGRGLG-----GGAGAAAAAAGAA 200
DB 316 GDAGDGGTGGRRGLAGAGGAGGAGGVTGGAGGAGFSRALIVAGDGDPPGAGAGGT 375
QY 201 GGAGQ--GGYGLGSGQTS-CRGLGGGAGGAGAAAAAAGGA-----GQGGYGLG 250
DB 376 GGAGSTTGAHGAAGASPTSGNGAGAGNGAHFSSGGKAGNGGAGGAGGLVNGGAGAG 435
QY 251 SQGT-----SRRGLGGGAGAAAAAAGGA--GQGGYGLGSGQTSGR 295
DB 436 GNGAPGPPSGGDPNNGGGGAGGAGGKGGDGAQAGDGGAGGKGGNGNGATGATGL 495
QY 296 GGL-----GGGAGAAAAAAGGA-----AGGA-----GQGGY 322
DB 496 NGLGAGADGTDGKGGNGAGGAGGGGAGGQGGKALAAATHDQDSMGAGGNGGAGMGD 555
QY 323 GGLGSGQT-----SGPGYGPQQTSGRGLGGGGA-----GAGGAGGAGG 353
DB 556 GNGAKGTFTDNGGDCVGGNGNGSGRIGGAGGLGGAGSTAGADGARGATPTSGNGGTG 615
QY 354 --GAAAAAAGAGAGGQGGYGL--GSQTSRGLGGGAGGAGAAAAAAGGAGQ--- 407
DB 616 GNGANATVAGGAGGAGKGGNGLVNGGAGGKGGDGMAGVAGSPPTAGESGTSGQNG 675
QY 408 -GGYGGGLGSQ-----GTSRGLGGGAGAAAAAAGGA-----AGQGGY 448
DB 676 AGGAGGAGGGRGDPGDPGTGGAGNGANGANATTPKAKGDDGGHGGPAGGNGGCGGP 735
QY 449 GGL-----GSGQTSRGLGGGAGAAAAAAGGAGGAGGAGGAGGAGGAGGAGG 491
DB 736 GGLAGNLFQNGIGVGGSGKGGAGGLAGDGGNANGNFAGDNGHGGHNGNPNPAGG 795
QY 492 QGTSPGGYGPQO-----TSGRGLGGGAGAAAAAAGGAGGQGGYGL--GSQ 542
DB 796 QGSGGAGSTPCAKGAHCFPTSGDGGDGGNGN-----SQVYGGNGDGGNGGSSA 850
```

```
QY 543 GTSRGRGLGGGAGAGAAAAA-----AAAGGAGGQGGYGLGSGQTSGRGG 587
DB 851 GTGNGRGGGDFAGFGMSANATNPENPNPGNGGAGGAGGAGLNG--GNGGAGNGG 909
QY 588 LCGGAGAAAAAAGAGGAGQ-----GGYGLG-----SQTSGRGLGGQ 629
DB 910 LGFGGNGCAAGANGVAVGAPQPGGAGHGGAGNGGAGGCGQVSDGAGGAGGAGD 969
QY 630 G-----AGAAAAAAGAGGAGGQGG--LGSQGTSGPGYGPQQTSGIRPAAKL 680
DB 970 GGAPGDGANGCGCAGAFAGGGGGRGDDGNAGNAGAGPGGTG---STAGKAGPAGSI 1026
```

RESULT 11

A70896  
hypothetical glycine-rich protein Rv1091 - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 24-Nov-1999  
C:Accession: A70896  
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome A: Reference number: A70500; MUID: 98295987; PMID: 9634230  
A: Accession: A70896  
A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
A: Molecule type: DNA  
A: Residues: 1-853 <COL>  
A: Cross-references: GB:AL021897; GB:AL123456; NID: g3256022; PIDN: CAAL1707.1; PID: e125  
A: Experimental source: strain H37RV  
C: Genetics:  
A: Gene: Rv1091  
C: Superfamily: unassigned collagens

Query Match 32.6%; Score 1139; DB 2; Length 853;  
Best Local Similarity 40.3%; Pred. No. 4.1e-57;  
Matches 298; Conservative 27; Mismatches 304; Indels 110; Gaps 28;

```
QY 11 GSMASGRGLGGGAGAAAA-----AAAAAGAGGQGGYGLGSGQTS-----GRGGL 58
DB 137 GGLLYNGNGNGTSTTAGVAGNGGAAGLNGGAGGGGAGAGGNGAGGWLTYNGGA 196
QY 59 GGQAGAAAAAAGGA-----GQGGYGLGSGQTS-----RGGLGGGAGAAAAA 107
DB 197 GGAGTSTVPGVAGNGGAGGAGSAGLWGTGGAGDGGNGRSGPVNVAGSAGNGAGGAAG 256
QY 108 AAAAGAGGQGGYGLG-----SQTSGRGLGGGAGAAAAAAGGAGGQGGY 159
DB 257 LFGDAGAGNGGKGGAGGAASFINFATAGDGGAGGSGGHALLWAGAGGNGGSGGTG 316
QY 160 GLG--SQTSPGGYGPQQT-----GRGLGGGGA-----GAAAAAAGGAGGQGGY 209
DB 317 GAGGTAGAGNGAGGAGGGGTGGLLFGNGAGAGHGAAGNGLAAGNGVSSGCGGAGGTG 376
QY 210 GLGSGQTSGRGLGGQ-----AGAAAAAAGGAGGQGGYGL-----GSQTSGR 258
DB 377 GAGDGG--GAGCAGGNARLWGVGAGGAGDGGAGGAGGKGGCGLSGNANGAGGDSRG 434
QY 259 GLGQGGAGAAAAAAGGAGGQGGYGLGSGQTSRGLGGGAGAAAAAAGGAG 318
DB 435 GTG--GAGGEGGAAGLLVGTGGHGGDGGAG--GAAVKGGDGAAGTGTAGAGGRRGAG 490
QY 319 QGGYGLGSGQTSPPGGYGPQQTSGRGLGGGAGAAAAAAGGAGGAGGAGGAGGAGG 371
DB 491 SGGSGDGGCGGAAGPAGW-----LFGDGGAGNGGAAAGGAGGAGGCGGNGCGNGG 545
QY 372 YGGLSGQTS-----GRGLGGGAGAAAAA-----AAGGAGGQGGYGLGSGQTSGR 422
DB 546 NGNGGNGATGWLTYNGGAGGQATAGAGGAGAGVSTNGGTGGNGGIG--CTGGSG 603
```

QY 423 GLGGQA--GAAAAAAGAG--GGYGGG----SQTSRGGGLGG--GAGAAAAA 474  
DB 604 GAGNAGLLGVGAGHGASGGADRGAGGTGFISSDGGAGDGGDGGNGGAGTGLL 663  
QY 475 AAAAGGAGGGYGLGSQTSRGGYGGPGQTSRGGYGGGAGAGAAAAAAGAGAG 534  
DB 664 FGAGNGGPGGSGAADIG--GNGGAGNGGCTDNGNGSGGG--AGSGGGGAGGN 718  
QY 535 GYGLGSQTSRGGYGGGAGAGAAAAAAGAGAGGGYGGGLGSQTSRGGYGGGAG 594  
DB 719 GAWLFGNGGAGGGGKGGAG--GGLGGGSGFLPGLNGSGDGDGNGAP 768  
QY 595 AAAAAAAGAGAGGGYGLGSQTSRGGYGGGAGAGAAAAAAGAGAGGGYGGG 654  
DB 769 GGYLYNGGAG--GQSSSGGIGGPGAT--GGAGKGGDGDGAQLIGDGGNGGAGGTG 824  
QY 655 SQTSRGGYGGPGQTSGI 673  
DB 825 --GTPGPGGSGGLGGL 841  
RESULT 12  
A44112  
spidroin 2, dragline silk fibroin - orb spider (Nephila clavipes) (fragment)  
N:Alternate names: silk fibroin, dragline  
C:Species: Nephila clavipes  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 20-Mar-1998  
C:Accession: A44112; S27824  
R:Hinman, M.B.; Lewis, R.V.  
J. Biol. Chem. 267, 19320-19324, 1992  
A:Title: Isolation of a clone encoding a second dragline silk fibroin. Nephila clavipes  
A:Reference number: A44112; MUID:92406876; PMID:1527052  
A:Accession: A44112  
A:Status: Preliminary  
A:Molecule type: mRNA  
A:Residues: 1-627 <HIN>  
A:Cross-references: GB:M92913; NID:g159713; PID:g159714  
R:Hinman, M.B.; Lewis, R.V.  
submitted to the EMBL Data Library, May 1992  
A:Description: Isolation of a clone encoding a second dragline silk fibroin: Nephila clavipes  
A:Reference number: S27824  
A:Accession: S27824  
A:Molecule type: mRNA  
A:Residues: 19-627 <H12>  
A:Cross-references: EMBL:M92913

Query Match 32.5%; Score 1135; DB 2; Length 627;  
Best Local Similarity 47.5%; Pred. No. 5.5e-57;  
Matches 324; Conservative 32; Mismatches 180; Indels 146; Gaps 36;  
QY 21 GGGAGAAAAAAGAGAGGGYGGGLGSQTSRGGYGGGAGAGAAAAAAG--G 76  
DB 2 GGYGPGQ-----GQPGYGP-GQGGPSGP-----SAAAAAAGAGGGY 42  
QY 77 AGGGYGGGLG-SQTSRGGYGGGAGAGAAAAAAGAGAGGGYGGGLGSQ-----GTSR 131  
DB 43 PGQQPGGYGPGQQPGRYGPGQQPGSPGSAAAAAA-GSGQQPGYGPQQPGPGY 101  
QY 132 GGLGGGAGAAAAAAGAGAGGGYGGGLGSQTSRGGYGGPGQTSRGGGL-----185  
DB 102 GQGGPSGSAASAAASAESGQQPGYGP-PGQQPGYGPQ--GPGYGPQQQPG 158  
QY 186 -GGAGAAAAAAGAGAGGGYGGGLGSQTSRGG--GLGGGAGAGAAAAAAGG 240  
DB 159 SGPSAAAAAAGAGAGGGYGGYGP-GQQPGYGPQQGPGSGPSAAAAAAGSGPG 217  
QY 241 AGGGYGGGLGSQTSRGGGLG-GQ-----GAGAAAAAAGAGAGGGYGGGLGSQTSR 295  
DB 218 QGPGYGP----PGQQPGYGPQQGLSGPSAAAAAAGPGQQPGYGP-GQQGPGSG 272  
QY 296 GGLGGGAGAAAAAAGAGAGGGYGGGLGSQTSRGGYGGPGQTSRGGYGGGAG 355

DB 273 G-----SAAAAAAGPGGYG-----PQQGPGYGPQ--GPGSAGS 310  
QY 356 AAAAAAAGAGAGGGYGGGLGSQTSRGG--GLGGGAGAGAAAAAAG--GAGGGY 410  
DB 311 AAAAAAGPGQGLGYP-GQQPGYGPQGGYGGYGGYGGYGGYGGYGGYGGYGGY 369  
QY 411 GGLGSQTSRGGYGGGAGAGAAAAAAGAGAGGGYGGGLGSQTSRGGYGGGAG 468  
DB 370 GP-GQQGPGSG-----SASAAAAAAGPGGYGP-GQQGPGYAPGQGGPSGP 417  
QY 469 AAAAAAAGAGAGGGYGGGLGSQTSRGGYGGPGYGGPGYGGPGYGGPGYGGPGY 528  
DB 418 SASAAAAA--AGPGYGP-----PQQGPGYAPGQ-----GPGSGPSAAAAA 463  
QY 529 GAGAGGGYGGGLGSQTSRGGYGGGAGAAAAAAGAGAGGGYGGGLGSQTSRGG 588  
DB 464 --AGPGYGP-AQQGPGSG-----IASASAGPGGYG-----PQQGP 500  
QY 589 GQCGAGAAAAAAGAGAGGGYGGGLGSQTSRGGYGGGAGAGAAAAAAGAGAG 648  
DB 501 AGYGPSSAASA---GAGSAGYGP-GSQASAAASRLASPDGSGARVASVNLVSSGPT 555  
QY 649 GYGGGLGSQ-----GTSRGP 662  
DB 556 SSAALSSVISNAVSYQICASNP 577  
RESULT 13  
A70812  
hypothetical glycine-rich protein RV0833 - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: A70812  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: A70812  
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-749 <COL>  
A:Cross-references: GB:AL022004; GB:AL123456; NID:g3261550; PIDN:CAAL17639.1; PID:g291  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: RV0833  
C:Superfamily: elastin  
Query Match 31.9%; Score 1114.5; DB 2; Length 749;  
Best Local Similarity 39.6%; Pred. No. 8.7e-56;  
Matches 303; Conservative 18; Mismatches 309; Indels 135; Gaps 32;  
QY 16 GRGGLGGCAGAAAAAAGAGAGGGYGGGLGSQ-----GTSR 55  
DB 3 GNGGAGSGAPGAGGAGGPGAGLIGVGGAGGDSAVAGVIGGAGGAGGAGGAGG 62  
QY 56 GGLGGGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 108  
DB 63 GCAGGSG-GSAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 121  
QY 109 AAAAGAGGGYGGGLGSQTSRGGYGGGAGAGAAAAAAGAGAGGGYGGGLGSQTS 167  
DB 122 GGTGGAGSGTGGVG--GTGGAGGLFASGGAGGSGGTGGAGTGGAGGLFAGGAG 179  
QY 168 GPGYGPQQTSGRGLGQ-----GAGAAAAAAGAGAGAGAGAGAGAGAGAGAG 215  
DB 180 GLG--GQNHGTGGHGGAGGAGGAGLALDGGAGGAGGAGGAGGAGGAGGAGG 237  
QY 216 TSGRGLGQ-----GAGAAAAAAGAGAGGGYGGGLGSQTSRGGYGGGAGAG 267  
DB 238 AGSGGAGTFTDGTNSGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 297



QY 11 GSMASRGGLGGGAGNAAAAAAGGAGCGGCGGGLGSGQTSGRGGLGGGAGAAAAA 70  
Db 139 GWWLLNGCNGSGCAPQQTGCAGGAALLGHGCTGGAG--GTGASGCKGCTGGLWMSGG 196  
QY 71 AAAAGGAGCGGCGGLGSGQ-----GTSRGLGG--QGAGAAAAAAGGA--CGGCGG 122  
Db 197 AGGAGSG--CGSGAGGNALMFIGNGGAGGAASGVNGVGAGGAGGAGGALVAIGGAG 255  
QY 123 LGSQGTSGRGLGGGAGAAAAAAGGAGCGGCGGGLGSGQTSRPGYG----PQQQT 178  
Db 256 AGCAATTGTCGAGGAGSN-ALGLFLGLGSGGCGGDSAMCGGAGGAGSGGAASPFID 314  
QY 179 SGRGGLGGGAGAAAAAAGAG-------QGGYGLSQGTSRGLGGG 226  
Db 315 IGIAGGCHG-GAGTNGGAGGAGGSGTVPFALDLWSGAGGNGCAATTGTGGAG--G 371  
QY 227 AGAAAAAAGGAGCGGCGGLG--SQGTSRGLGGGAGAAAA--AAAAAGGA 279  
Db 372 TGGFNAVDFIGFAYGGAGGLGGATGAGTGGTGGVGGGFAALGVGVGGAGGAGGA 431  
QY 280 -----QGGYGLGSGQTSRGLGGGAGAAAAAAG-----GAGGCGGL 325  
Db 432 ATETGGTGGAGGLGVGLGCGAGGAGGPG-CAASAGSGHGTGDDALGLGAGTGGVGV 490  
QY 326 G-----SQGTSRPGGCGPQQTSGRGLGGGAGAAAAAAGGAGCGGCGGLSQGT 380  
Db 491 GCAATDTGNGGAGGSGTG--LLGGVCGAGGHHGG--ASVGTGGSGGAGGDFGVGAGGN 547  
QY 381 SGRGGLG-----GQAGAAAAAAGGAGCGGCGGLGSGQTSRGLGGGAGAAAA 434  
Db 548 GGNAGTGVGVNGANGNGSNTGALAAVGA-----GAAGDATSGTGGFG--GAGGSAR 600  
QY 435 AAAAAAGGAGCGGCGGLGSGQTSRGLGGGAGAAAAAAGGAG--CGGYG----- 487  
Db 601 GLIFALGGAGAA--GGDASTGVGPGGPGCTGTASSPFGIAIAGGAGAGGAGTSCATG 658  
QY 488 -----GLGSGQTSRPGGCGPQ-QTSRGLGGGAGAAAAAAGGAG 532  
Db 659 GAGGCGVFEGIAVLGLGFGAGAGGAATGDGATGGAGFGCAGAGTANFLGFSVLHG-G 717  
QY 533 QGGYGLGSGQTSRGLGGGAGAAAAAAGGAGCGGCGGLGSGQTSRGLGGG 592  
Db 718 AGGAGTAT-GTGGNGAGGGGGLSPVILIGIGGAGGDDGALGVLG--GMGGDGGDG 774  
QY 593 AGAA-----AAAAAAGGAGCGGCGGLGSGQTSRGG--GLGCGAGAAAAAAGGAG 646  
Db 775 GEAVAVGIAVGGAGGAGGAAPTCTGCGAGGNGGDALGLVGVGNGGAGTCTFGANTGNGG 834  
QY 647 QG-----GYGGLGSGQTSRPGGCGP 667  
Db 835 DTTIVVNGMLAPSTLGYGNGGNGVNGGAG-GTG 867

Search completed: December 18, 2002, 16:22:06  
Job time : 29.1216 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:06:04 ; Search time 10.4528 Seconds  
(without alignments)  
2702.183 Million cell updates/sec

Title: US-09-490-291-4

Perfect score: 3493

Sequence: 1 MRGSHHHHHGSMASRGGL.....GGYGPQQQTSGIRRRPAKLN 681

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2135	61.1	747	1 SPDL_NEPCL	P19837 nephila cla
2	1277.5	36.6	1901	1 YZ08_MYCTU	O53553 mycobacteri
3	1168	33.4	5263	1 FBOH_BOMMO	O57590 bombyx mori
4	1135	32.5	627	1 SPDL_NEPCL	P46804 nephila cla
5	1106.5	31.7	914	1 WA22_MYCTU	O06794 mycobacteri
6	1044.5	29.9	778	1 YQ34_MYCTU	P19333 mycobacteri
7	1028	29.4	957	1 Y278_MYCTU	P56877 mycobacteri
8	988	28.3	801	1 Y747_MYCTU	O53810 mycobacteri
9	976	27.9	860	1 ELS_MOUSE	P54320 mus musculu
10	939	26.9	864	1 ELS_RAT	O99372 rattus norv
11	869	24.9	603	1 YD25_MYCTU	Q10637 mycobacteri
12	853.5	24.4	465	1 GRP2_PHAVU	P10496 phaseolus v
13	775.5	22.2	747	1 ELS_BOVIN	P04985 bos taurus
14	763	21.8	641	1 EBN1_EBV	P03211 epstein-bar
15	748.5	21.4	1356	1 CA21_ONCMY	O93484 oncorhynch
16	704.5	20.2	750	1 ELS_CHICK	P07916 gallus gall
17	698.5	20.0	1355	1 CA21_RANCA	O42350 rana catesb
18	691	19.8	730	1 ELS_HUMAN	P15502 homo sapien
19	686	19.6	498	1 Y118_MYCTU	O50615 mycobacteri
20	679	19.4	1733	1 VNVA_PRVKA	P33485 pseudorabie
21	678	19.4	543	1 YP91_MYCTU	O50630 mycobacteri
22	673	19.3	384	1 GRP1_PETHY	P09789 petunia hyb
23	669	19.2	481	1 LOR1_MOUSE	P18185 mus musculu
24	655	18.8	1027	1 CAFF_RIFPA	P30754 riftia pach
25	644	18.4	338	1 GRP_ARATH	P27483 arabidopsis
26	626.5	17.9	1262	1 CA13_CHICK	P12105 gallus gall
27	626	17.9	491	1 YK98_MYCTU	Q10707 mycobacteri
28	626	17.9	515	1 Y140_MYCTU	Q50594 mycobacteri
29	625	17.9	1372	1 CA21_MOUSE	Q01149 mus musculu
30	621	17.8	1466	1 CA13_HUMAN	P02461 homo sapien
31	608	17.4	1372	1 CA21_RAT	P02466 rattus norv
32	608	17.4	1453	1 CA11_CHICK	P02457 gallus gall
33	607.5	17.4	1418	1 CA12_HUMAN	P02458 homo sapien

#### ALIGNMENTS

RESULT 1

ID	SPDL_NEPCL	STANDARD;	PRT;	747 AA.
AC	P19837;			
DT	01-FEB-1991 (Rel. 17, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Spidroin 1 (Dragline silk fibroin 1) (Fragment).			
OS	Nephila clavipes (Orb spider).			
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;			
OC	Araneomorphae; Entelegynae; Araneolidae; Tetragnathidae; Nephila.			
OX	NCBI_TaxID=6915;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE=90384959; PubMed=2402494;			
RA	Xu M., Lewis R.V.;			
RT	"Structure of a protein superfiber: spider dragline silk.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).			
RN	[2]			
RP	SEQUENCE OF 653-747 FROM N.A.			
RX	MEDLINE=94165058; PubMed=8120021;			
RA	Beckwith R., Arcidiacono S.;			
RT	"Sequence conservation in the C-terminal region of spider silk proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and Araneus bicentenarius (Araneidae).";			
RL	J. Biol. Chem. 269:6661-6663(1994).			
CC	-!- FUNCTION: Spiders major ampullate silk possesses unique characteristics of strength and elasticity. Fibroin consists of pseudocrystalline regions of antiparallel beta-sheet interspersed with elastic amorphous segments.			
CC	-!- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 2, OF THE DRAGLINE SILK.			
CC	-!- SUBCELLULAR LOCATION: Extracellular.			
CC	-!- DOMAIN: Highly repetitive protein characterized by regions of polyalanine and glycine-rich repeating units.			
CC	-!- SIMILARITY: BELONGS TO THE SILK FIBROIN FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; M37137; AAA29380.1; -;			
DR	EMBL; U03848; AAB60212.1; -;			
DR	PIR; A36068; A36068.			
KW	Silk; Repeat.			
FT	NON_TER	1	655	25 X APPROXIMATE TANDEM REPEATS.
FT	DOMAIN	1	25	1.
FT	REPEAT	1	25	2.
FT	REPEAT	26	38	3.
FT	REPEAT	39	66	4.
FT	REPEAT	67	96	5.
FT	REPEAT	97	130	6.
FT	REPEAT	131	158	

```
FT REPEAT 159 191 7.
FT REPEAT 192 204 8.
FT REPEAT 205 235 9.
FT REPEAT 236 262 10.
FT REPEAT 263 292 11.
FT REPEAT 293 305 12.
FT REPEAT 306 333 13.
FT REPEAT 334 360 14.
FT REPEAT 361 394 15.
FT REPEAT 395 424 16.
FT REPEAT 425 458 17.
FT REPEAT 459 485 18.
FT REPEAT 486 512 19.
FT REPEAT 513 525 20.
FT REPEAT 526 555 21.
FT REPEAT 556 582 22.
FT REPEAT 583 612 23.
FT REPEAT 613 642 24.
FT REPEAT 643 655 25.
FT CONFLICT 662 682 V -> L (IN REF. 1).
FT CONFLICT 672 S -> T (IN REF. 1).
FT CONFLICT 695 NFGLSGCDVLIQALLVWSALIQILGSSSIGOVNYSAGQA
FT TQILGQSVYQAL -> ILVFLDVMSSFKLFLRLLSRS
FT (IN REF. 1).
SQ SEQUENCE 747 AA; 60528 MW; 850E44B0D649E012 CRC64;

Query Match 61.1%; Score 2135; DB 1; Length 747;
Best Local Similarity 68.3%; Pred. No. 7.9e-89;
Matches 495; Conservative 15; Mismatches 39; Indels 176; Gaps 34;

QY 23 QGAGAAAAAAGAGAGAGGGYGLGSGQT--SGRGLGGGQAG-AAAAAAGAGAGG 79
DB 1 QGAG-----AAAAAGAGAGGGYGLGGQAGGGYGLGGQAGAGAGAAAAAGAGG 56
QY 80 GYGGLGSGQTSRGLGGGAGAGAAAAAAGAGAGGGYGLGSGQTSRGLGGGAGA 139
DB 57 GYGGLGSGQ-----AGRGGQAG-----AAAAAGAGAGGGYGLGGQAG 107
QY 140 GAAAAAAGAGAGAGGGYGLGSGQTSRGGYGPQQTSGRGLGGGAGAGAAAAA 199
DB 108 G-----AAAAAGAGGGYGLGNQ-----AGRGGQ-----AAAA 142
QY 200 AGAGAGGGYGLGSGQTSRGLGGGAGAGAAAAA-----AGAGAGGGYGLG 250
DB 143 AGAGAGGGYGLGSGQ-AGRGLGGGAGAGAAAAAGAGGGYGLGGGAGAGGGYGLG 201
QY 251 SQGTSRGLGGGAGAGAAAA-----AAAAAGAGGGYGLGSG 289
DB 202 SQG-AGRGLGGGAGAGAAAAAGAGAGGGYGLGGGAGAGAGAGGGYGLGSG 260
QY 290 QGTSRGLGGGAGAGAAAAAAGAGAGGGYGLGSGQTSRGGYGP-GQQTSGRGL 348
DB 261 QG-----AGRGGEGAG-----AAAAAGAGGGYGLGGQ-AGQGGYGLGSGAGRGL 311
QY 349 GGGAGAGAA-----AAAAAGAGAGGGYGLGSGQTSRGLGGG 390
DB 312 GGGAGAGAGAGGGYGLGGGAGAGAGAGAGAGGGYGLGSGQ-AGRGLGGG 370
QY 391 AGAAAAAAGAGAGGGYGLGSGQTSRGLGGGAGAGAGAGAGAGAGAGAGAGAGAG 450
DB 371 AG-----AVAAAAAGAGAGGGYGLGSGQ-----AGRGGQAG-----AAAAAGAGAGGGY 419
QY 451 LGSQTSRGLGGGAGAGAGAGAGAGAGAGGGYGLGSGQTSRGLGGG-----SGP 497
DB 420 LGNQ-AGRGLGGGAG--AAAAAAGAGAGGGYGLGNQAGRGGAGAGAGAGAGAG 475
QY 498 GYGP-GQQTSGRGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 550
DB 476 GYGGLGSGQAGR--GGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 531
QY 551 GGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 590
DB 532 GGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 587
```

```
QY 591 QGAGAAAAAAGAGAGGGYGLGSGQTSRGLGGGAGAGAGAGAGAGAGAGAGAG 650
DB 588 QGAG-----AAAAAGAGAGGGYGLGSGQV-GRGLGGGAG-----AAAGAGAGGGY 636
QY 651 GGLGS 655
DB 637 GVGGS 641

RESULT 2
YZ08_MYCTU STANDARD; PRT; 1901 AA.
AC O53553;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PE-PGRS family protein RV3508 precursor.
GN RV3508 OR MTV023.15.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Horsley R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Dornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AL022022; CAAL17745.1; -.
DR TubercuList; RV3508; -.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00534; PE; 1.
DR ProDom; PD001223; PE_region; 1.
KW Hypothetical protein; Repeat; Signal; Complete proteome.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1901 HYPOTHEICAL PE-PGRS FAMILY PROTEIN
FT RV3508.
SQ SEQUENCE 1901 AA; 147627 MW; C7B1923D5D0146CD CRC64;

Query Match 36.6%; Score 1277.5; DB 1; Length 1901;
Best Local Similarity 39.7%; Pred. No. 1.1e-50;
Matches 351; Conservative 28; Mismatches 270; Indels 235; Gaps 39;

QY 11 GSWASRGGLGQ-GAGAAAAAAGAGAGGGYGLGSGQ-----GTS 53
DB 292 GGALGTGTGTGTGGAGGAGGRCALLLLGAGGGGLGAGGGGTGGAGDGLGVGGTG 351
QY 54 GRGLGG-----QGAGAAA-----AAAAAAGG-AGGGYGLGSGQ----- 89
DB 352 GKGGVGVAGLGGAGGACQLFSAGGAGAVGGTGGGGAGGAGAGADAPASTGLTG 411
QY 90 ----TSGRGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 136
```









Qy	593	-----AGAAAAAAAAAGGACQGGYGG--LGSQGTSTGRGGTGGQGAGAAAAAAAAAGGAGQ	647
Db	822	RSFEGGAGGAGGASDGTGAGGCGAGGKAGLYNGGDBGAGGAGTSGKGGAGGNAVVICN	881
Qy	648	GCYGGGLSQ--GTSGPGGYG	665
Db	882	GGGGNAGKAGGTAGAGGAG	901

RESULT 6	YQ34_MYCTU	YQ34_MYCTU	STANDARD;	PRT;	778 AA.
ID	YQ34_MYCTU				
AC	P71933;				
DT	01-NOV-1997	(Rel. 35, Created)			
DT	01-NOV-1997	(Rel. 35, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	DE	Hypochemical PE-PCRS family protein RV2634C.			
GN	RV2634C OR MT2712	OR MTCY441.04C.			
OS	MYcobacterium	tuberculosis.			
OC	Bacteria;	Actinobacteria;	Actinobacteria (class);	Actinobacteridae;	
OC	Actinomycetales;	Corynebacterineae;	Mycobacteriaceae;	Mycobacterium.	
OX	NCBI_TaxID=1773;				
RN	[1]				
RP	SEQUENCE FROM N.A.				

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR	EMBL; Z802235; CAB02341.1; -.
DR	EMBL; AE007103; AAK47026.1; ALT_INIT.
DR	TIGR; MT2712; -.
DR	Tuberculinist; RV2634c; -.
DR	InterPro; IPR000084; PE_region.
DR	Pfam; PF0934; PE; 1.
DR	ProDom; PD001223; PE_region; 1.
KW	Hypothetical protein; Complete proteome.
FT	CONFLICT 51 51 V -> L (IN REF. 2).
FT	CONFLICT 63 63 Q -> H (IN REF. 2).
FT	CONFLICT 274 274 A -> T (IN REF. 2).
SQ	SEQUENCE 778 AA; 63131 MW; DAB20FE584999E7 CRC64;

RESULT 7			
ID	Y278_MYCTU	STANDARD;	PRT; 957 AA.
AC	P56877;		
DT	30-MAY-2000 (Rel. 39, Created)		
DF	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Hypothetical PS-PGRS family protein RV0278c precursor.		
GN	RV0278C OR MT0291 OR MTV035.06C.		
OS	Mycobacterium tuberculosis.		
OC	Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;		
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.		
OX	NCBI_TaxID=1773;		
RX	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=H37Rv;		
RX	MEDLINE=98295987; PubMed=9634230;		
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,		
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,		
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,		
RA	Davies R., Devlin K., Felwell T., Gentles S., Hamlin N, Holroyd S.,		

[illegible]







QY 296 GGLGGGAGAGAAAAAGGAGGGYGGGLG--SQGTSGPGYGPQGTSGRGLGG--Q 351  
 Db 413 GIVGPGCAVSPAAAKAAKAAKAYGARGVGIIPTYG-VAGGF-PGYVGAGAGLGGASQ 470  
 QY 352 GAGAAAAAAGGAGGGYGGGL-----GSGQGTSGRG 384  
 Db 471 AAAAAAANKYAGAGAGTGLGLVPGAVPGALPGAVPGALPGAVPGVPGTG 530  
 QY 385 GLGGGAGAGAAAAA-----AGGAGGGY-GGLGSGQTSG----- 420  
 Db 531 GVPAGTTPAAAAAATAAATAAAGAGYGLGPGVGVGGVGLPGVPGVGTGIGTGP 590  
 QY 421 -----RGLGGGAGAAAAA-----AAAG-GAGGGY-----GLGSGQTS 457  
 Db 591 GTGLVPGDLGGAGTAAAKAAKAAKAAKAYRAAAGLGCAGVPLGVGAGVPGFAGAGGFG 650  
 QY 458 GRGLGGGQAG-----AAAAAAGGAGGGYGGGLGSGQTSGRGLGGGAGAAAAAGGA 511  
 Db 651 AGAVPGFAGAVPGSLAKAAKAYNAGLGGPLGGPLGGPLGGPGGFG-----GPGG 704  
 QY 512 LGGQAGAA--AAAAAAGGAGGGYGGGLGSGQTSGRGLGGGAGAAAAAAGGA 569  
 Db 705 LGGVPGVAGGAPAAAAAATAAQAQYGLGAGGLGAGGLGAGGLG-----AGGL 755  
 QY 570 GGGY--GGLGSGT--SGRGLGGGAGAAAAAAGGAGGGYGGGLGSGQTSGRGL 626  
 Db 756 GAGGLGAGGLGAGGVTPGAVGV--GVSPAAAAAKAAKAYGAAGLGGV--LGAFFPGGVA 811  
 QY 627 GGGGAGAAAAAAGGAGGAGG-----YGG-LGSGTSGPGGYG 665  
 Db 812 ARPGFGLSPYGGGAGGLGVGKPKPYGGALCALGYGGGCGF 856

RESULT 11  
 ID YD25\_MYCTU STANDARD; PRT; 603 AA.  
 AC Q10637;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical PE-PGRS family protein Rv1325c precursor.  
 GN Rv1325C OR MT1367 OR MTCY130.10C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37Rv;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fieischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Uitterback T., Weidman J., Khouri J., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS

CC SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z73902; CAA98089.1; --  
 DR EMBL: A5007010; AAK45630.1; ALT\_INIT.  
 DR TIGR: MT1367; --  
 DR Tuberculist; Rv1325c; --  
 DR InterPro; IPR000084; PE\_region.  
 DR Pfam; PF00934; PE; 1.  
 DR ProDom; PD001223; PE\_region; 1.  
 DR Hypothetical protein; Repeat; Signal; Complete proteome.  
 KW SIGNAL 1 30  
 FT CHAIN 31 603  
 FT DOMAIN 114 603  
 FT CONFLICT 132 132 G -> V (IN REF. 2).  
 FT CONFLICT 135 135 G -> D (IN REF. 2).  
 FT CONFLICT 337 337 G -> D (IN REF. 2).  
 FT CONFLICT 508 508 L -> F (IN REF. 2).  
 SQ SEQUENCE 603 AA; 49575 MW; 4F9BCB82B07AE964 CRC64;  
 Query Match 24.9%; Score 869; DB 1; Length 603;  
 Best Local Similarity 38.0%; Pred. No. 7e-33;  
 Matches 256; Conservative 35; Mismatches 256; Indels 126; Gaps 29;  
 QY 26 GAAAAAAGGAGGGYGGGLGSGTSGRGLGGGAGAAAAAAGGAGGGYGGGL 85  
 Db 22 GSTLGAANAALCP-----TTLLAAGADEVSAAIASLFAHQQA----YQAV 65  
 QY 86 GSQGTSGRGL-----GGQGAGA-AANAAGAGGAGGGYGGGLGSGTSGRGLGGGAGA 141  
 Db 66 SAQMSAFHAQFVQTETAGAYASAEAAAAAPLEGLLNIVNTPTQLLLRPLNGANGA 125  
 QY 142 AAAAAAAGGA--GGGYGGGLGSG--TSCPG--YGPQGTSGRGLGGGAGAAAA 194  
 Db 126 PTGQAGGAGGLLYNGGAGGSGAPGAGPAGGANGLFNG-----GAGGAGGDPGNGA 181  
 QY 195 AAAAAAGG--GGGYGGGLGSGTSGRGLGGGAGAAAAAAGGAGGGYGGGLGSG 252  
 Db 182 GGAGGAGLLFSGGAGGPGGVNTGTGGLGGDG--GAAGLFGAGGIGGAGGPGFNG---- 236  
 QY 253 GTSRGLGGGAGAAAAAAGGAG-----GGYGGGLGSGTSGRGLG---GGQAGA 305  
 Db 237 ---GAGGAGR-SGLFEVLAAGGAGGTGGLSVNGTGGTG--CTGGGGGLFSGGAGGAG 290  
 QY 306 AAAAAAAGGAGGGYGGGL--GSGTSGPGYGPQ--TSRGLGGGAGAAAAAAGGA 362  
 Db 291 GFVSGSAGNGGTGGDGGIFTGNGTGTGTGNTQLVGGEGGAGGAGNAGILFGAG 350  
 QY 363 AAGGAGGGYGGGLGSGTSGRGLGGGAGAAAAAAGGAGGGYGGGLG--SQGTSGR 421  
 Db 351 GIGTGTGLGAPDPGPGTGGKGVGGIG-----GAGALFGPGCAGTGGFGCASSADQWA 404  
 QY 422 GGLGGGAGAAAAAAGGAGGGYGGGLGSGTSGRGLGGGAGAAAAAAGGA 481  
 Db 405 GGIGSG--GSGGAALKLGDGAGGTGGDSVRGAAGSGGTGCTG-----GLI 449  
 QY 482 GGGYGGGLGSGTSGPGYGPQGTSGRGLGGGAGAAAAAAGGAGGGYGGGLG 541  
 Db 450 GGGAGGAG-----GTGIEFSGVGGAGGAGNAAAGLSGAGGAGGA--GGFG---- 493  
 QY 542 QGTSGRGLGGGAGAAAAAAGGAGGGYGGGLGSGTSGRGLGGGAGAAAAAAGGA 601  
 Db 494 -ETAGDGGAGN-----AGLLNDGGAGGAGGLGIAGDGNKGKGGK----- 534  
 QY 602 AAGGAGGGYGGGLGSGTSGRGLGGGAGAAAAAAGGAGGGYGGGLG--QGTSG 660









```

FT CARBOHYD 1257 1257 (BY SIMILARITY).
SQ SEQUENCE 1356 AA; 126985 MW; 7BB2F1F80DB10C93 CRC64;
  Query Match 21.4%; Score 748.5; DB 1; Length 1356;
  Best Local Similarity 31.6%; Pred. No. 2.3e-27;
  Matches 253; Conservative 44; Mismatches 368; Indels 135; Gaps 29;

QY 15 SGRGLGGGQAGAAAAAAGAGQGGY-GGLGSGQTSG-RGGLGGQAGAAAAAAA 72
Db 108 TGHAGEPGEPTGSGIGARGPTGSACKPGEDGNNRPGKPGDGGGPTQGA-RGFPCTPG 166

QY 73 AAGGACGGYGGI-GSQGTSRGGLGCGCAGAAAAAAGAG--QGGYGGGSGQTS 130
Db 167 LPMKGHRYNGLDGRKSGTAGAKGTGAHGANGPGPAGSRGLNGERAGPAGPAG 226

QY 131 RGLGGGQAGAAAAAAGAG--QGGYGGGSGQTSGP-GGYG-PGQQ---- 177
Db 227 ARGADSTGPAGPLGAAGPPGFPAGPGKEIGGAGSNGSPGQGGRGEPGINGAVG 286

QY 178 -----TSRGGGLGGQAGAAAAAAGAG--QGGYGGGSGQTSRGGLGGQAGAA 230
Db 287 PVPVGNPNNGINGAKGAAGLPVAGAFPGFPGPRGPGPGQPGQSTGARGLGSDPSPG 346

QY 231 AAAAAAAGCAGCGG-YGGLGSGQTSRGGLGGQAGAAAAAAGAGCGGCGG-- 287
Db 347 QKGDGSAKGEPHSGHYGQGAAGPAGEEGKRGSTGVGTGATGALRGAGGAGTRGLPLEG 406

QY 288 -----GSQGTSRGGLGGQAGAAAAA-----AAAAAGGAGGG-----YGGLG 326
Db 407 RGPPTMPARATATGCGINGAPDAGRAGESGLTGARGLPNSGGGPPGKEGPPGAAG 466

QY 327 SQGTSG-PGGYGPQQ-----TSRGGGLGGQAGAAAAAAGAGCGGCGG-- 375
Db 467 LDGRTGPPGPTGPRGQPNIGFPKPGPGGKAGKDGKPTGATGLRGPGADGNNGAPG 526

QY 376 -----GSQGTSRGGLGGQAGAAAAAAGAGCGGCGG-----GSQG 417
Db 527 PAVVNTGTEKEGQGPAGAPGQGLPGPAPAGEAGKAGNQGNMPGQGLPGPAGVKGERG 586

QY 418 TSG-RGGLGGQAGAAAAAAGAGGOGGYGGLGSGQTSRGGLGGQAGAAAAAAA 476
Db 587 NSGPAGSAGSQAIGARGPACTPDPDGKGPESVIVGAAGHQGGMPGERGAGGTPG 646

QY 477 AAGGAGQGGYGGI-----GSQGTSGP-----GGYGPQQTSGRGLG 513
Db 647 PRKEKEGGHRLGEGNMGRDARGGPGSPGPPSGANGKEKESGSGFPAGPAGLRGPSS 706

QY 514 --GQAGAAAAAAGAGCGG-----YGGGSGQTSRGGLGGQAGAAAAAAA 564
Db 707 ERGEGGAGLPAGPAGPPGSDGSGPRGKGPAGGKGDVGPAGPAGPSPGSPGASGPAG 766

QY 565 AAGGAGQGGYGGI-GSQGTSRGGLGGQAGAAAAAAG-----GAGOG 610
Db 767 PFGGRGDAGPSGLTFPPGAAGRVGGPGPACIAGPPGACPKDPRGLRDPGPGGPG 826

QY 611 GYGGLGSGQTSRGGLGGQAGAAAAAAGAGCGGCGG--GSQGTSG-PGGYG--- 665
Db 827 EQGVVGPAGISGDKGPSGESGPPGAPGTAGPQVLPSPGFVGLPGSRGDKGLPGGPGAVG 886

QY 666 -PGQ-----QTSGIRPAAKL 680
Db 887 EFGRLGPAGASGPRGPAGNI 906
```

Search completed: December 18, 2002, 16:18:04  
Job time : 20.4528 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:13:04 ; Search time 39,982 Seconds  
(without alignments)  
3509.535 Million cell updates/sec

Title: US-09-490-291-4

Perfect score: 3493

Sequence: 1 MRGSHHHHHGSMASRGGL.....GCGPGQQPSGIRPPAAKLN 681

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2009	57.5	644	5 Q8WSW4	Q8WSW4 nephila cla
2	1991.5	57.0	617	5 O46172	O46172 nephila cla
3	1726.5	49.4	544	5 O46171	O46171 nephila cla
4	1592	45.6	691	5 Q9BIU3	Q9BIU3 dolomedes t
5	1568	44.9	648	5 Q9BIU7	Q9BIU7 argiope tri
6	1531.5	43.8	854	5 Q9BIU4	Q9BIU4 dolomedes t
7	1450	41.5	988	5 O17434	O17434 nephila cla
8	1448	41.5	1953	5 Q9BIT7	Q9BIT7 nephila mad
9	1386.5	39.7	2639	5 O76786	O76786 antheraea p
10	1350	38.6	2655	5 Q964F4	Q964F4 antheraea y
11	1267	36.3	1217	16 Q8VIY9	Q8VIY9 mycobacteri
12	1263.5	36.2	1715	16 Q8VI20	Q8VI20 mycobacteri
13	1254.5	35.9	1489	16 O53559	O53559 mycobacteri
14	1228.5	35.2	1079	16 O53557	O53557 mycobacteri
15	1199.5	34.3	447	5 Q9BIV1	Q9BIV1 argiope aur
16	1197	34.3	1306	16 O53775	O53775 mycobacteri

17	1186.5	34.0	1384	16 Q8VIZ1	Q8VIZ1 mycobacteri
18	1182.5	33.9	1381	16 O53552	O53552 mycobacteri
19	1173	33.6	1685	16 O53215	O53215 mycobacteri
20	1160.5	33.2	1507	16 Q8VJ23	Q8VJ23 mycobacteri
21	1159.5	33.2	1884	5 Q9NHW2	Q9NHW2 nephila mad
22	1158	33.2	1002	5 Q9BIU8	Q9BIU8 argiope tri
23	1157	33.1	738	5 O02402	O02402 pinctada fu
24	1157	33.1	1538	16 O53395	O53395 mycobacteri
25	1152.5	33.0	1408	16 Q8VK17	Q8VK17 mycobacteri
26	1151	33.0	2249	5 Q9NHW4	Q9NHW4 nephila cla
27	1148	32.9	1468	5 Q9GUB5	Q9GUB5 galliera me
28	1145	32.8	1329	16 O06810	O06810 mycobacteri
29	1139	32.6	853	16 O53439	O53439 mycobacteri
30	1114.5	31.9	749	16 O53844	O53844 mycobacteri
31	1099	31.5	907	5 O44359	O44359 nephila cla
32	1083	31.0	871	5 O44358	O44358 nephila cla
33	1080.5	30.9	879	16 Q8VKD2	Q8VKD2 mycobacteri
34	1075	30.8	882	16 O53845	O53845 mycobacteri
35	1041.5	29.8	563	5 Q9BIT5	Q9BIT5 nephila mad
36	1036	29.7	767	16 O53435	O53435 mycobacteri
37	1024	29.3	837	16 O53684	O53684 mycobacteri
38	1015.5	29.1	651	5 Q9BIU9	Q9BIU9 argiope tri
39	1014	29.0	783	16 O53809	O53809 mycobacteri
40	1011	28.9	636	5 Q16987	Q16987 araneus dia
41	1005	28.8	775	16 Q8VJ15	Q8VJ15 mycobacteri
42	1004	28.7	731	16 O50415	O50415 mycobacteri
43	984.5	28.2	714	16 O53556	O53556 mycobacteri
44	984	28.2	741	16 O06808	O06808 mycobacteri
45	975	27.9	810	11 Q9ES29	Q9ES29 mus musculu

#### ALIGNMENTS

RESULT 1

Q8WSW4 ID Q8WSW4 PRELIMINARY; PRT; 644 AA.  
AC Q8WSW4;  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Dragline silk protein (Fragment).  
OS Nephila clavipes (Orb spider).  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;  
OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.  
OX NCBI\_TaxID=6915;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Basel R.;  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY061814; AAL32375.1; -;  
DR InterPro; IPR00104; Antifreeze\_1.  
DR PRINTS; PR00308; ANTIFREEZE1.  
FT NON\_TER 644  
SQ SEQUENCE 644 AA; 52104 MW; 9024AF2F65C5F75A CRC64;

Query Match	57.5%	Score	2009;	DB 5;	Length	644;
Best Local Similarity	70.9%	Pred. No.	7.9e-101;			
Matches	461;	Conservative	12;	Mismatches	39;	Indels 138; Gaps 30;
QY	30	AAAAAAGGAGGGYGGGLGSGQTSGRGLGGGAGAAAAAAGGAGGGYGGGLGSGQ	89			
Db	2	AAAAAAGGAGGGYGGGLGSGQ-----AGRGGGAG-----AAAAAAGGAGGGYGGGLGSGQ	53			
QY	90	TSRGGGLGGGAGAGAAA-----AAAAAAGGAGGGYGGGLGSGQT	128			
Db	54	AGRGLGGGAGAGAAAAAGGVCQGLGGGAGAGAAAAAGGAGGGYGGGLGSGQ-	111			
QY	129	SGRGLGGGAGAGAAAAAAGGAGGGYGGGLGSGQTSGPGYGP-QQTSRGGGLGGQ	187			
Db	112	---AGRGGGAGAGAAAAAV---GAGGGYGGQ-----AGGGYGGGLGSGQAGRGLGGQ	160			
QY	188	GAGAAAAAAGGAGGGYGGGLGSGQTSGRGLGGGAGAGAAAAAAGGAGGGYGGY	247			









```
QY 583 SGRGLGGGAGAAAAAAGAG-GAGGCGYGGGLGSGQTSRGLGGGAGAAAAA 641
Db 691 -GGG-Y-GGSGGAGAAVAASGAGGAGGGYGGGAGGAGGAGGAGGAGGAG 744
QY 642 AGGAGGCGYGGGLGSGQTSRGGYGGGAGGAGGAGGAGGAGGAGGAGGAG 733
Db 745 AGGAG-GGGGCGYGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 775

RESULT 7
O17434
ID O17434 PRELIMINARY; PRT; 988 AA.
AC O17434;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE Minor ampullate silk protein Misp1 (Fragment).
OS Nephila clavipes (Orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
OX NCBI_TaxID=6915;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98200471; PubMed=9541398;
RA Colgin M.A., Lewis R.V.;
RT "Spider minor ampullate silk proteins contain new repetitive sequences
RT and highly conserved non-silk-like 'spacer regions'.";
RL Protein Sci. 7:667-672(1998).
DR EMBL; AF027735; AAC14589.1; -.
DR InterPro; IPR000817; Prion.
DR PRINTS; PR00341; PRION.
FT NON_TER 1
FT SEQUENCE 988 AA; 79082 MW; 461E03DF53F7085D CRC64;
SQ

Query Match 41.5%; Score 1450; DB 5; Length 988;
Best Local Similarity 50.6%; Pred. No. 9.3e-71;
Matches 354; Conservative 17; Mismatches 212; Indels 117; Gaps 25;

QY 21 GGGGAGAAAAAAGAGGAGGCGYGGGLGSGQTSRGLGGGAGAAAAAAGGAGG 80
Db 220 GGYGAG---AGAVAAAGAGAGGCGY---RGAGGCGGCGYGGAGAGAGAAAA--GAGAG 270
QY 81 GYGLGSGQTSRGLGGGAGAAAAAAGAGGAGGCGY---GLGSGQTSRGLG 135
Db 271 GAGGY-----GRGAGAGAAAAAGAGAGGAGGAGGCGYGGGAGAGAAAAAGAGGAG 324
QY 136 GGGGAGAAAAAAGAGGAGGCGYGGGLGSGQTSRGGYGGGAGGAGAAAAA 195
Db 325 GYRGAGAGAGAAA--GAGAGGCGGCGYGGAGAGAGAAAAAGAGGAGGAGAGAG 382
QY 196 AAAAGAGGCGYGGGLGSGQTSRGLGGGAGAAAAAAGAGGAGGCGYGGGLGSGQTS 255
Db 383 GAAAGAGAGAGSY-----GGGCGYGGAGAGAAAAAGAGAGAGSY----- 422
QY 256 GRGLGGGAGAAAAA--AAAAAGAGGCGYGGGLG-----SQGTSRGLGGGAGGAGAA 306
Db 423 GRGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 482
QY 307 AAAAAAGAGGAGGCGYGGGLGSGQTSRGGYGGGAGGAGGAGGAGGAGGAGGAGGAG 363
Db 483 AGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 542
QY 364 AGGAGGCGYGGGLGSGQTS-----SGRGLGGGAGAAAAAAGAGGAGGCGYGG 412
Db 543 AAGAGAGGCGYGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
QY 413 LGSGQ-----TSRGLGGGAGAAAAAAGAGGAGGCGYGGGLGSGQTSRGG 461
Db 601 QGGYGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 651
QY 462 LGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 517
```

```
Db 652 QGGYGGAGAGAGAAAAAGAGAGGAGYGRGAGAGAGAGAGAGAGAGAGAGAGG 711
QY 518 GAAAAAAGAGAGGCGYGGGLGSGQTSRGLGGGAGAAAAAAGAGAGGCGYGG-- 575
Db 712 GAGAGAGAGAGAGAGAGAGAGY-SRG--GRAGAGAGAGAGAGAGAGAGAGAGG 768
QY 576 -GLGS-----QGTSGRGLG-GGAGAGAGAGAGAGAGAGAGAGAGAGAGG 628
Db 769 AGAGAGAGAGAGGAGGAGYGRGAGAGAGAGAGAGAGAGAGAGAGAGAGG-- 817
QY 629 QGAGAGAGAGAGAGAGGCGYGG-GLGSGQTSRGGYGGG 667
Db 818 YGAGAGAGAGAGAG--GAGRGGYGRGAGAGAGGCGYGGG 855

RESULT 8
Q9BIT7
ID Q9BIT7 PRELIMINARY; PRT; 1953 AA.
AC Q9BIT7;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE Major ampullate spidroin 2-like protein (Fragment).
OS Nephila madagascariensis.
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
OX NCBI_TaxID=115969;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21179804; PubMed=11283372;
RA Gatesy J., Hayashi C., Motriuk D., Woods J., Lewis R.;
RT "Extreme Diversity, Conservation, and Convergence of Spider Silk
RT Fibroin Sequences.";
RL Science 291:2603-2605(2001).
DR EMBL; AF350276; AAK30605.1; -.
FT NON_TER 1
FT NON_TER 1953
FT SEQUENCE 1953 AA; 159383 MW; C82B4DACD043C3BD CRC64;
SQ

Query Match 41.5%; Score 1448; DB 5; Length 1953;
Best Local Similarity 38.3%; Pred. No. 2e-70;
Matches 451; Conservative 25; Mismatches 173; Indels 530; Gaps 51;

QY 11 GSWASRGGGLGGGAGAGAAAA-----AAAAAGAGGCGYGGGLGSGQTSRGG 57
Db 546 GGYGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 598
QY 58 LGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 83
Db 599 PGGPGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 658
QY 84 GLGSGQ-----TSRGG-----GLGCGGAGAGAGAGAGAGAGAGAGAGAG 114
Db 659 GPGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 718
QY 115 AGGGYGG-----GLGSGQTSRGG-----GLGCGGAGAGAGAGAGAGAGAG 144
Db 719 SGPGYGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 778
QY 145 -AAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 169
Db 779 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 838
QY 170 GYGCGGQ-----TSRGG-----GLGCGGAGAGAGAGAGAGAGAGAGAGAG 194
Db 839 GYGCGGQGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 898
QY 195 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 232
Db 899 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 958
QY 233 -----AAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 266
```

Db 959 GPGGYPGQOQPGCAAAAAAGSGPGYGP-GQQGPGPGCAAAAAAGRGPGGYGPQOQGP 1017  
QY 267 AAAAAAAGAGAGGGY-----GLSQGTSGRG-----GLGQAG 304  
Db 1018 GPGAAAAAA-GRGPGYGPQOQPGQOQPGSGAAAAAAGRGPGGYGPQOQPGPGAA 1076  
QY 305 AAA-----AAAAAGAGAGGGY-----GLSQGT'S-----GPGYGP 339  
Db 1077 AAAAGPGYGPQOQPGCAAAAAAGRGPGGYGPQOQPGPGCAAAAAAAGRGPGGYGP 1136  
QY 340 QOTSGRGLGGGAGAAAA-----AAAAAGAGGGYGLGSGQTSGRGL 386  
Db 1137 QOQPGQOQPGPGCAAAAAAGRGPGGYGPQOQGTCAAAAAAGSAGGY-----PGQGP 1189  
QY 387 GQAGAAAA-----AAAAAGAGGGY-----G 412  
Db 1190 GPGAAAAAAGPGYGPQOQPGCAAAAAAGSGPGYGPQOQPGSGSSAAAAAAGPRYG 1249  
QY 413 LGSQGT-----SGRGL--GGGAGAAAAAAGAGAGGGYGLGSGQTSGRG- 460  
Db 1250 PQOQPGAAAAAAGSGPGYGPQOQPGPGCAAAAAAAGRGPGGYGP-GQQGPGPGA 1308  
QY 461 -----GLGGGAGAAAA-----AAAAAGAGGGY-- 487  
Db 1309 AAAAGPGYGPQOQPGCAAAAAAGSGPGYGPQOQPGPGCAAAAAAGRGPGGYGP 1368  
QY 488 --GLSQG-----TSRGGYGPQOQ-----TSRGL--GGGAGAAAAAAG 529  
Db 1369 QOQPGPGAAAAAAGPGYGPQOQPGCAAAAAAGSGPGYGPQOQPGCAAAAAA 1428  
QY 530 GAGGGY-----GLSQG-----TSRG-----GLGG 552  
Db 1429 GRPGYGPQOQPGPGCAAAAAAGPGYGPQOQPGCAAAAAAGRGPGGYGPQOQPGG 1488  
QY 553 QAGAAAA-----AAAAAGAGGGYGLGSGQTSRG----- 586  
Db 1489 SCAAAAAGRGPYGPQOQPGPGCAAAAAAGTGPYGP-GQQPGCAAAAAAGR 1547  
QY 587 -----GLGGGAGAAAA-----AAAAAGAGGGYGLGSGQTSGR 623  
Db 1548 GPGYGPQOQPGPGCAAAAAAGPGYGPQOQPGCAAAAAAGSGPGYGP-GQQPGPG 1606  
QY 624 G-----GLGGGAGAAAA-----AAAAAGAG 647  
Db 1607 GAAAAAGRGPYGPQOQPGPGCAAAAAAGRGPYGPQOQPGSGNAAGSGRGP 1666  
QY 648 GGYG-----GLSQG-----TSRGGYGPQOQTSIRPAA 678  
Db 1667 GYGPQOQPGPGCAAAAAAGPGYGPQOQPGCAAAAA 1705  
RESULT 9  
O76786 PRELIMINARY; PRT; 2639 AA.  
ID O76786  
AC O76786;  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE Fibroin.  
OS Antherea pernyi (Chinese oak silk moth).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
OC Bombycoidea; Saturniidae; Saturniinae; Saturniini; Antherea.  
OX NCBI\_TaxID=7119;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20496953; PubMed=11040284;  
RA Sezutsu H., Yukihiko K.;  
RT "Dynamic rearrangement within the antherea pernyi silk fibroin gene  
is associated with four types of repetitive units.";  
RL J. Mol. Evol. 51:329-338(2000).  
DR EMBL: AF083334; AAC32605.1;  
SQ SEQUENCE 2639 AA; 216057 MW; 2EE3310DEEB09B9A CRC64;

Query Match 39.7%; Score 1386.5; DB 5; Length 2639;  
Best Local Similarity 40.4%; Pred. No. 5e-67;  
Matches 402; Conserved 23; Mismatches 239; Indels 330; Gaps 34;  
QY 2 RGRHHHHCMA-----SGRGLG-----QAGAAAAAAGAGAGGGY 45  
Db 135 RGGFYETHSYSSYSGSSAAAGAGAGGGYCGSDSAAAAAAGAGGGY 194  
QY 46 GLGSO-----GTSRGLGGGAGAAAAAAGAGAGGGYGG 86  
Db 195 GYGSDSAAAAAAGAGAGGGYGGYCGSDSAAAAAAGAGAGGGY 254  
QY 87 SQGTSGRGLGGGAGAAAAAAGAGAGGGYGLGSGQTS----- 129  
Db 255 GYGWGDGGYGSAAAAAAGAGAGGGYGSAAAAAAGAGAGGGY 314  
QY 130 -----GRGLGGGAGAAAAAAGAGAGGGYGLGSO----- 164  
Db 315 AGGAGGGYCGDGGYCGSDSAAAAAAGAGAGGGYCGSDSAAAAAAGAGAGGGY 374  
QY 165 -----GTSRGLGGGAGAAAAAAGAGAGGGYCGSDSAAAAAAGAGAGGGY 201  
Db 375 ARRACHDRAGSAAAAAAGAGAGGGYCGW-GDGGYGSAAAAAAGAGAGGGY 433  
QY 202 GAGGGYGG-----LGSQTSRGLGGGAGAAAAAAGAGAGGGY 235  
Db 434 GSGAGAGGGYCGWDSGSDSAAAAAAGAGAGGGYCGYGSAAAAAAGAGAGGGY 493  
QY 236 AAAGAGAGGGYGLGSGQTSRGLGGGAGAAAAAAGAGAGGGYGLGSG 291  
Db 494 AAAGAGAGGGY--GSYGW-GDGGYCGSDSAAAAAAGAGAGGGYCGSDSAAAAAAGAGAGGGY 550  
QY 292 T-----SGRGLGGGAGAAAAAAGAGAGGGY--GGY-----GGLGSGGT--- 330  
Db 551 AAAAAAASARRACHDRAGSAAAAAAGAGAGGGYCGWGGYGSAAAAAAGAGAGGGY 610  
QY 331 -----SGPGYGPQOQTSRGLGGGAGAAAAAAGAGAGGGY--GGY-----GGLG 452  
Db 611 AAAAAAAGAGAGGGYGW-GDGGYGSAAAAAAGAGAGGGYCGYGSAAAAAAGAGAGGGY 669  
QY 379 GT-----SGRGLG-----QAGAAAAAAGAGAGGGY 410  
Db 670 SAAAAAAGAGAGGGYCGWGGYCGSDSAAAAAAGAGAGGGYCGY 729  
QY 411 GGLGSGGT-----SGRGLGGGAGAAAAAAGAGAGGGY--GGY-----GGLG 452  
Db 730 GSGSSAAAAAASARRACHDRAGSAAAAAAGAGAGGGYCGWGGY 789  
QY 453 SGT-----SGRGLG-----QAGAAAAAAGAGAGGGY 486  
Db 790 SDSAAAAAAGAGAGGGYCGWGGYGSAAAAAAGAGAGGGYCGY 849  
QY 487 GGLGSGGT-----SGPGYGPQOQTSRGLGGGAGAAAAAAGAGAGGGY--AG 529  
Db 850 GYGSDSAAAAAAGAGAGGGYCGW-GDGGYGSAAAAAAGAGAGGGY 908  
QY 530 GAGGGYGLGSGGT-----SGRGLGGGAGAAAAAAGAGAGGGY--GGY 575  
Db 909 GRGDGGYGSAAAAAASARRACHDRAGSAAAAAAGAGAGGGYCGY 968  
QY 576 GLGSGQTSRGLGGGAGAAAAAAGAGAGGGYCG----- 614  
Db 969 W-----GDGGYGSAAAAAAGAGAGGGYCGWGGYGSAAAAAAGAGAGGGY 1021  
QY 615 -----LGSQTSRGLGGGAGAAAAAAGAGAGGGY-----GGLGSO----- 656  
Db 1022 AAAAAAGAGGGYCGYGSAAAAAAGAGAGGGYCGWGGYCGSDSAAAAAAGAGAGGGY 1081  
QY 657 -----GTSRGLGGGAGAAAAAAGAGAGGGYCGY 678  
Db 1082 AAAAAAGAGAGGGYCGWGGYGSAAAAAAGAGAGGGY 1115





```
QY 665 GPGQQTSG 672
Db 1054 DGGGGGQ 1061

RESULT 13
O53559 ID O53559 PRELIMINARY; PRT; 1489 AA.
AC O53559;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE PGRS-family protein.
GN RV3514 OR MV023.21.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
DR EMBL; AL022022; CAA17751.1;
DR TuberculList; RV3514;
DR InterPro; IPR000084; PF_kinase.
DR DR
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE_kinase; 1.
DR PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN_2.
DR Complete proteome.
KW
SQ SEQUENCE 1489 AA; 115754 MW; 6855CBA1C3CBAF3A CRC64;

Query Match 35.9%; Score 1254.5; DB 16; Length 1489;
Best Local Similarity 39.5%; Pred. No. 3.7e-60;
Matches 336; Conservative 36; Mismatches 267; Indels 211; Gaps 36;

QY 16 GRGGLGGG-----AGAAAAAAGAGGAGQ-----GGYGLGSGQTSGRGGLGGGAGGA 65
Db 260 GHGVGGDGWLAPCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 317

QY 66 AAAAAAAGGA-QGGYGGGLGSGQ-----GTSGRGLGQ-----QGAGAAAAAAGGA 113
Db 318 LGAGGQGLGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 377

QY 114 GAGGQGGYGLGSGQTSGRGGLGGGGA-----GAAAAAAGGA-----GQGG 157
Db 378 AAGNAGVGGAGGQG--GDGGAGGAGAGADAPGATGCTGTFAGGAGGAGGAGGAGGAGGAGG 435

QY 158 YGGLGSGQTSQPGYGPQQTSGRGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 216
Db 436 SGGAGGQGGAGGAGGAGADNPITGIGTGGDGGTGAAGAGGAGGAGGAGGAGGAGGAGGAGG 495

QY 217 SGRGLGAGG-----GAGAAA-----AAAAAAGGA-----GQGGYGG 248
Db 496 AGVGAGGQGGGAGGAGGAGADAPGATGCTGTFAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 555

QY 249 LGSQG-----TSGRGGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 295
Db 556 TGGGQGGAGGAGGAGADNPITGIGTGGDGGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 615

QY 296 GGLGGQ----GAGAAAAA-----NAAAGGAGQ-----GQYGGGLGS 327
```

Query Match

35.2%; Score 1228.5; DB 16; Length 1079;



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:05:10 ; Search time 46.1374 Seconds  
(without alignments)  
1995.696 Million cell updates/sec

Title: US-09-490-291-6  
Perfect score: 3543  
Sequence: 1 MASWTGQGMGRGSMASGRG.....TSGIRIRAPSTFEHHHHH 691

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3543	100.0	691	22	AA82610 Spider recombinant
2	3390	95.7	681	22	AA82609 Spider recombinant
3	2431	68.6	1255	23	AA50037 N. clavipes spidro
4	2431	68.6	1880	23	AA50039 N. clavipes spidro
5	2384	67.3	989	23	AA50038 N. clavipes spidro
6	2321.5	65.5	1626	23	AAU11798 Dragline protein 1
7	2298	64.9	1617	23	AAU11794 Dragline protein 1
8	2283.5	64.5	818	23	AAU11797 Dragline protein 1
9	2254	63.6	809	23	AAU11793 Dragline protein 1
10	2151	60.7	630	23	AA50042 N. clavipes spidro

11	2151	60.7	676	23	AA50047 N. clavipes spidro
12	2138	60.3	718	19	AA53346 Nephila clavipes s
13	2138	60.3	718	21	AA53070 N. clavipes spider
14	2135	60.3	651	20	AA40097 Spider silk protei
15	2135	60.3	651	23	AAU11781 Spider natural sil
16	2129	60.1	718	12	AA14308 N.clavipes draglin
17	2076	58.6	606	16	AA99053 Spider dragline va
18	2076	58.6	606	20	AA40100 Polymer of an anal
19	2059	58.1	646	18	AA27178 Nephila clavipes s
20	1972.5	55.7	606	16	AA99055 Spider dragline va
21	1972.5	55.7	606	20	AA40102 Polymer of an anal
22	1971.5	55.6	604	16	AA99057 Spider dragline va
23	1963.5	55.4	606	20	AA40101 Polymer of an anal
24	1784.5	50.4	528	22	AA82611 Spider recombinant
25	1669.5	47.1	777	23	AA50046 N. clavipes spidro
26	1614	45.6	773	23	AA50045 N. clavipes spidro
27	1458.5	41.2	364	23	AA50043 N. clavipes spidro
28	1405.5	39.7	831	16	AA80168 pMIS1 MISP spider
29	1400	39.5	615	20	AA40099 Spider silk protei
30	1392.5	39.3	655	23	AB876672 Protein related to
31	1357	38.3	231	12	AA82608 Spider recombinant
32	1346.5	38.0	714	16	AA99059 Spider dragline va
33	1346.5	38.0	714	20	AA40103 Polymer of an anal
34	1241.5	35.0	1177	9	AA80940 SLP111 protein com
35	1241.5	35.0	1177	11	AA805307 SLP III (Silk-fibr
36	1241.5	35.0	1177	17	AA95105 Silk like protein
37	1241.5	35.0	1177	18	AA26342 SLP111 amino acid
38	1241.5	35.0	1177	21	AA78277 Repetitive protein
39	1241.5	35.0	1177	22	AA72725 SLP111 protein seq
40	1241.5	35.0	1177	22	AA863995 Silk-like protein
41	1241.5	35.0	1178	14	AA41007 SLP111-like sequenc
42	1241.5	35.0	1178	19	AA53518 Amino acid sequenc
43	1184.5	33.4	1023	19	AA53524 SLP4 multimeric pr
44	1184.5	33.4	1059	14	AA41013 SLP4 synthetic pro
45	1184.5	33.4	1059	18	AA26348

ALIGNMENTS

RESULT 1  
AA82610  
ID AA82610 standard; Protein; 691 AA.  
XX  
AC AA82610;  
XX  
XX 02-OCT-2001 (first entry)  
XX  
XX Spider recombinant silk protein pET((SP1)4/(SP2)1)4.  
DE  
DE Spider; orb-weaver; silk protein; pET((SP1)4/(SP2)1)4;  
KW structural protein; purification; fibre; spinning.  
XX  
OS Nephila clavipes.  
XX  
XX WO200153333-A1.  
PN  
XX 26-JUL-2001.  
PD  
XX 01-NOV-2000; 2000WO-US30086.  
PF  
XX 20-JAN-2000; 2000US-0490291.  
PR  
XX (MELL/) MELLO C M.  
PA (ARCI/) ARCIDIAcono S.  
PA (BUTL/) BUTLER M M.  
PA (USSA ) US SEC OF ARMY.  
XX  
PI Mello CM, Arcidiacono S, Butler MM;  
XX WPI; 2001-483136/52.  
DR N-PSDB; AA826303.  
XX

PT Recovering structural polypeptides in a biological sample, useful for  
PT purifying and spinning spider silks and other structural proteins,  
PT comprises treating the sample containing the polypeptides with an acid  
PT \_  
XX  
XX  
XX  
PS Claim 2; Page 38-40; 49pp; English.  
PS  
CC The present sequence is that of orb-weaver spider (Nephila clavipes)  
CC recombinant silk protein pET((SP1)4/(SP2)1)4. The invention  
CC provides methods for purifying and spinning spider silks and other  
CC structural proteins. Organic acids are used to lyse recombinant  
CC cells or other biological samples (such as non-recombinant  
CC derived cells), and enrich the purity and yields of structural  
CC proteins by hydrolysing many of the macromolecules while leaving  
CC the structural proteins intact. In the case of silk proteins, the  
CC resulting lysate is further purified by ion-exchange or affinity  
CC chromatography and processed into an aqueous-based mixture for  
CC fibre spinning. In the present case, the pET((SP1)4/(SP2)1)4 gene  
CC was cloned into vector pET24 for recombinant expression in  
CC Escherichia coli, and recombinant silk protein was obtained in  
CC 75-75% purity using propionic acid and anion-exchange chromatography  
CC on QAE-Sephadex A50. When denaturant (3 M guanidine-HCl) was added  
CC to the propionic acid, the recombinant protein was obtained in 80%  
CC purity. Products obtained using the methods of the invention can  
CC be used in the construction of many materials including films,  
CC fibres, woven articles, sutures, ballistic protection, parachutes  
CC and parachute cords. The new method has the following advantages  
CC over prior art: it involves fewer steps, requires less time and  
CC smaller volumes of reagents, results in better recovery of protein  
CC at higher purity (70-99%), is easy to scale up, and the fibres are  
CC spun in an environmentally benign solution reducing hazardous waste  
CC accumulation and cost.  
XX  
SQ Sequence 691 AA:  
  
Query Match 100.0%; Score 3543; DB 22; Length 691;  
Best Local Similarity 100.0%; Pred. No. 6.3e-246; Indels 0; Gaps 0;  
Matches 691; Conservative 0; Mismatches 0;  
  
QY 1 MASMTGGQMGSGMASGRGLGGQAGAGAAAAAGAGGAGGQGGYGLGSQGTSGRGL 60  
DB 1 MASMTGGQMGSGMASGRGLGGQAGAGAAAAAGAGGAGGQGGYGLGSQGTSGRGL 60  
  
QY 61 GGGGAGAAAAAGAGAGGAGGQGGYGLGSQGTSGRGLGGQAGAGAAAAAGAGGAG 120  
DB 61 GGGGAGAAAAAGAGAGGAGGQGGYGLGSQGTSGRGLGGQAGAGAAAAAGAGGAG 120  
  
QY 121 GYGGGLGSQGTSGRGLGGQAGAGAAAAAGAGGAGGQGGYGLGSQGTSGPGGYGPGQQT 180  
DB 121 GYGGGLGSQGTSGRGLGGQAGAGAAAAAGAGGAGGQGGYGLGSQGTSGPGGYGPGQQT 180  
  
QY 181 SGRGGLGGQAGAGAAAAAGAGGAGGQGGYGLGSQGTSGRGLGGQAGAGAAAAAGAG 240  
DB 181 SGRGGLGGQAGAGAAAAAGAGGAGGQGGYGLGSQGTSGRGLGGQAGAGAAAAAGAG 240  
  
QY 241 GGAGGQGGYGLGSQGTSGRGLGGQAGAGAAAAAGAGGAGGQGGYGLGSQGTSGRGL 300  
DB 241 GGAGGQGGYGLGSQGTSGRGLGGQAGAGAAAAAGAGGAGGQGGYGLGSQGTSGRGL 300  
  
QY 301 GGQAGAGAAAAAGAGGAGGQGGYGLGSQGTSGPGGYGPGQQTSGRGLGGQAGAGAA 360  
DB 301 GGQAGAGAAAAAGAGGAGGQGGYGLGSQGTSGPGGYGPGQQTSGRGLGGQAGAGAA 360  
  
QY 361 AAAAAAGAGGQGGYGLGSQGTSGRGLGGQAGAGAAAAAGAGGAGGQGGYGLGSQGT 420  
DB 361 AAAAAAGAGGQGGYGLGSQGTSGRGLGGQAGAGAAAAAGAGGAGGQGGYGLGSQGT 420  
  
QY 421 SGRGGLGGQAGAGAAAAAGAGGAGGQGGYGLGSQGTSGRGLGGQAGAGAAAAAGAG 480  
DB 421 SGRGGLGGQAGAGAAAAAGAGGAGGQGGYGLGSQGTSGRGLGGQAGAGAAAAAGAG 480  
  
QY 481 GGAGGQGGYGLGSQGTSGPGGYGPGQQTSGRGLGGQAGAGAAAAAGAGGAGGQGGY 540  
DB 481 GGAGGQGGYGLGSQGTSGPGGYGPGQQTSGRGLGGQAGAGAAAAAGAGGAGGQGGY 540

Db 481 GGAGGQGGYGLGSQGTSGPGGYGPGQQTSGRGLGGQAGAGAAAAAGAGGAGGQGGY 540  
QY 541 LGSQGTSGRGLGGQAGAGAAAAAGAGGAGGQGGYGLGSQGTSGRGLGGQAGAGAAA 600  
DB 541 LGSQGTSGRGLGGQAGAGAAAAAGAGGAGGQGGYGLGSQGTSGRGLGGQAGAGAAA 600  
QY 601 AAAAAAGAGGQGGYGLGSQGTSGRGLGGQAGAGAAAAAGAGGAGGQGGYGLGSQGT 660  
DB 601 AAAAAAGAGGQGGYGLGSQGTSGRGLGGQAGAGAAAAAGAGGAGGQGGYGLGSQGT 660  
QY 661 SGPGGYGPQQTSGIRIRAPSTSFHHHHHH 691  
DB 661 SGPGGYGPQQTSGIRIRAPSTSFHHHHHH 691  
  
RESULT 2  
AAB82609  
ID AAB82609 standard; Protein; 681 AA.  
XX  
AC AAB82609;  
DT 02-OCT-2001 (first entry)  
XX  
DE Spider recombinant silk protein pQE((SP1)4/(SP2)1)4.  
KW Spider; orb-weaver; silk protein; pQE((SP1)4/(SP2)1)4;  
KW structural protein; purification; fibre; spinning.  
XX  
OS Nephila clavipes.  
XX  
PN WO200153333-A1.  
XX  
PD 26-JUL-2001.  
PF 01-NOV-2000; 2000WO-US30086.  
XX  
PR 20-JAN-2000; 2000US-0490291.  
XX  
PA (MELL/) MELLO C M.  
PA (ARCI/) ARCIDIACONO S.  
PA (BUTL/) BUTLER M M.  
PA (USSA ) US SEC OF ARMY.  
XX  
PI Mello CM, Arcidiacono S, Butler MM;  
XX  
DR WPI: 2001-483136/52.  
DR N-PSDB; AAH26302.  
XX  
PT Recovering structural polypeptides in a biological sample, useful for  
PT purifying and spinning spider silks and other structural proteins,  
PT comprises treating the sample containing the polypeptides with an acid  
PT \_  
XX  
PS Claim 2; Page 34-37; 49pp; English.  
XX  
CC The present sequence is that of orb-weaver spider (Nephila clavipes)  
CC recombinant silk protein pQE((SP1)4/(SP2)1)4. The invention  
CC provides methods for purifying and spinning spider silks and other  
CC structural proteins. Organic acids are used to lyse recombinant  
CC cells or other biological samples (such as non-recombinant  
CC derived cells), and enrich the purity and yields of structural  
CC proteins by hydrolysing many of the macromolecules while leaving  
CC the structural proteins intact. In the case of silk proteins, the  
CC resulting lysate is further purified by ion-exchange or affinity  
CC chromatography and processed into an aqueous-based mixture for  
CC fibre spinning. In the present case, the pQE((SP1)4/(SP2)1)4 gene  
CC was cloned into vector pQE-9 for recombinant expression in  
CC Escherichia coli, and recombinant silk protein was obtained in 97%  
CC purity using propionic acid and anion-exchange chromatography on  
CC QAE-Sephadex A50. Products obtained using the methods of the  
CC invention can be used in the construction of many materials  
CC including films, fibres, woven articles, sutures, ballistic  
CC protection, parachutes and parachute cords. The new method has







```
Db 421 GAGAAAAAGAGGAGGGYGGGLGSGAGRGGQAGAGAAAAAGAGAGGGYGGGLGSGAGRGQ 480
QY 295 -----SRRGGLGGGAGAGAAA----- 310
Db 481 GAGAAAAAGAGAGGGYGGGLGSGAGGGYGGGLGSGAGRGGLGGGAGAGAAAAAGAGGAG 540
QY 311 -----AAAAAGAGGGYGGGLGSGGT-----SGPGGYGP- 340
Db 541 GGLGGGAGAGAGAGAGAGAGGGYGGGLGSGAGRGGAGAGAAAAAGAGAGGGYGGGL 600
QY 341 GQQTSGRGLGGGAGAGAAAAAGAGAGGGYGGGLGSGGT--SRRGGLGGGAGAGAAA 398
Db 601 GSGAGRGGLGGGAGGAG-----AAAAAGAGGGYGGGLGSGAGGGYGGGLGSGAGGAG 656
QY 399 AAAAAAGAGAGGGYGGGLGSGGT-----SRRGGLGGG 429
Db 657 AAAAAAGAGAGGGYGGGLGSGAGRGGAGAGAAAAAGAGAGGGYGGGLGSGAGRGGLGG 716
QY 430 GAGAAAA-----AAAAAGAGGGYGGGLGSGGTSGRG 463
Db 717 GAGAAAAAGAGAGGGYGGGLGSGAGRGGAGAGAAAAAGAGGGYGGGLGSGG-AGRG 775
QY 464 LGGAGAGAAAAAGAGAGGGYGGGLGSGGTSGPGGYGP-GQQTSGRGLGGGAGAGAA 522
Db 776 LGGAGG-----AAAAAGAGGGYGGGLGSGG-AGGGYGGGLGSGAGRGGLGGGAGAA 830
QY 523 AA-----AAAAAGAGGGYGGGLGSGGT----- 546
Db 831 AAAAAAGAGGGYGGGLGSGAGAGAGAAAAAGAGGGYGGGLGSGAGRGGAGAGAAAAAG 890
QY 547 -----SRRGGLGGGAGAGAAA----- 562
Db 891 AGGGYGGGLGGGAGAGGGYGGGLGSGAGRGGLGGGAGAGAAAAAGAGAGGGYGGGLGSGAGG 950
QY 563 --AAAAAGAGAGGGYGGGLGSGGTSGRGLGGGAGAGAAA----- 600
Db 951 AGAAAAAGAGAGGGYGGGLGSGG-AGRGGLGGGAGAGAAAAAGAGAGGGYGGGLGSGAGR 1009
QY 601 -----AAAAAGAGAGGGYGGGLGSGGTSGRGLGGGAGAGAAA----- 638
Db 1010 GGCGAGAGAAAAAGAGGGYGGGLGSGG-AGRGGLGGGAGAGAAAAAGAGAGGGYGGGLGSG 1068
QY 639 -----AAAAAGAGAGGGYGGGLGSGGTSGPGGYGP 669
Db 1069 GAGRGGAGAGAAAAAGAGGGYGGGLGSGG-AGRGGAGG 1108

RESULT 5
ID AAM50038
AC AAM50038 standard; Protein; 989 AA.
XX
XX AAM50038;
XX 18-SEP-2002 (first entry)
DT
DE N. clavipes spidroin synthetic homologue SO1SM12 protein.
XX
KW Spidroin; spider; silk; fibre; film; membrane; wound; filter; SO1SM12.
XX
OS Synthetic.
XX
PN DE10113781-Al.
XX
PD 13-DEC-2001.
XX
XX 21-MAR-2001; 2001DE-1013781.
XX
XX 09-JUN-2000; 2000DE-1028212.
XX
XX 24-OCT-2000; 2000DE-1053478.
XX
XX (IPKP-) IPK INST PFLANZENGENETIK & KULTURPFLANZE.
XX
XX Scheller J, Conrad U, Grosse F, Guehrs K;
```

```
XX
DR WPI; 2002-123561/17.
DR N-PSDB; ABL61039.
XX
PT New DNA encoding synthetic spider silk protein, useful e.g. for closing
PT wounds, comprises modules that encode repeating units of spidroin
PT proteins
XX
XX Claim 22; Page 34-38; 88pp; German.
XX
CC This invention describes a novel DNA sequence, encoding a synthetic
CC spider silk protein, comprising modules, each comprising a group of
CC sequentially arranged oligonucleotides, each oligonucleotide encoding
CC a repeating unit of a spidroin protein. The synthetic protein has at
CC least 84% homology with the Nephila clavipes spidroin protein and is used
CC to produce synthetic fibres, films and/or membranes, particularly: (i)
CC for medical use, especially to close wounds and/or to support or cover
CC artificial organs; (ii) as adhesion surfaces for culturing cells; and
CC (iii) as filters. The synthetic proteins are very similar to native
CC spider silk proteins; can be prepared on a large scale and can be spun to
CC fibres with excellent mechanical properties (strength and elasticity).
CC Also they retain water solubility after long-term boiling in aqueous
CC solutions and since they are also soluble in organic solvents but
CC precipitated at high salt concentration, they are easily extracted and
CC purified. The modular construction of the invention facilitates
CC incorporation of additional peptide-encoding sequences, e.g. to simplify
CC purification or modulate solubility. This sequence represents the
CC synthetic N. clavipes spidroin-1 homologue SO1SM12 described in the
CC invention.
XX
SQ Sequence 989 AA;
Query Match 67.3%; Score 2384; DB 23; Length 989;
Best Local Similarity 57.8%; Pred. No. 5.8e-163;
Matches 565; Conservative 18; Mismatches 38; Indels 356; Gaps 35;
QY 6 GGQWGRGSMASRRGLGGGAGAGAAAAAGAGAGGGYGGGLGSGGTSGRGLGGGAG 65
Db 9 GGQAGGG-----GYGGGGGAGAGAGAAAAAGAGAGGGYGGGLGSGG-----AGRGGAG 60
QY 66 GAAAAAGAGAGAGGGYGGGLGSGGTSGRGLGGGAGAGAAA----- 108
Db 61 G-----AAAAAGAGAGGGYGGGLGSGG-AGRGGLGGGAGAGAAAAAGAGAGGGYGGGLGSG 115
QY 109 -----AAAAAGAGAGGGYGGGLGSGGTSGRGLGGGAGAGAAA-----AA 198
Db 116 AGRGGGAGAGAAAAAGAGAGGGYGGGLGSGG-AGRGGLGGGAG-----AAAAAGAGAGGG 170
QY 160 YGGLGSGGTSGPGGYGP-GQQTSGRGLGGGAGAGAAA-----AA 198
Db 171 YGGLGGG-AGGGYGGGLGSGAGRGGLGGGAGAGAAAAAGAGAGGGYGGGLGSGAGAGAA 229
QY 199 AAAAGAGAGGGYGGGLGSGGT----- 218
Db 230 AAAAGAGAGGGYGGGLGSGAGRGGAGAGAAAAAGAGAGGGYGGGLGGAGAGGGYGGGLG 289
QY 219 --SRRGGLGGGAGAGAAA-----AAAAAGAGAGGGYGGGLGSGGT 256
Db 290 QGAGRGGLGGGAGAGAAAAAGAGAGGGYGGGLGSGGAGAGAAAAAGAGAGGGYGGGLGSGG- 348
QY 257 SRRGLGGGAGAGAAA-----AAAAAGAGAGGGYGGGLG 290
Db 349 AGRGLGGGAGAGAAAAAGAGAGGGYGGGLGSGAGRGGAGAGAAAAAGAGAGGGYGGGLG 408
QY 291 SQGTSGRGLGGGAGAGAAAAAGAGAGGGYGGGLGSGGTSGPGGYGPQQT'S----- 345
Db 409 SQG-AGRGGLGGGAG-----AAAAAGAGAGGGYGGGLGSGG-AGRGGAGAGAAAAAGAG 462
QY 346 ---GRGLGGGAG---GAAAAAGAGAGGGYGGGLGSGGT-----SCRGG 387
Db 463 GGGYGGGLGSGAGRGGAGAGAAAAAGAGAGGGYGGGLGGGAGAGGGYGGGLGSGAGAGGG 522
QY 388 LGGAGAGAAA-----AAAAAGAGAGGGYGGGLGSGGTSGRGLG 427
```

|||||  
523 LGGGAGAAAAAGGAGCGGLGGGAGCGAGAGAAAAAGAGCGGGLGSGG-----AGRG 578  
QY 428 GCGAGAAAAAAGAGAGCGGCGGGLGCGTSGRGLGGGAGAAAAAAGAGAGCGG 487  
Db 579 GCGAG-----AAAAAGAGCGGCGGGLGSGG-AGRGGLGGGAG-----AAAAAGAGCGG 629  
QY 488 YGGLGSGTSGPGGYP-GQOTSGRGLGGGAGAAAA-----AA 526  
Db 630 YGGLGSGG-AGCGGGLGSGGAGRGGLGGGAGAAAAAAGAGCGGGLGGGAGCGAGAA 688  
QY 527 AAAGAGCGGCGGGLGSGTSGRGLGGGAGAAAA-----562  
Db 689 AAAGAGCGGCGGGLGSGG-AGRGGLGGGAGAAAAAAGAGAGCGGCGGAGCGGCG 747  
QY 563 --AAAAAGAGCGGCGGGLGSGTSGRGLGGGAGAAAA-----600  
Db 748 AGAAAAAGAGCGGCGGGLGSGG-AGRGGLGGGAGAAAAAAGAGAGCGGCGGGLGSGGACR 806  
QY 601 -----AAAAAGAGCGGCGGGLGSGG-----622  
Db 807 GCGAGAAAAAAGAGCGGCGGGLGSGGAGRGGCGGAGAAAAAAGAGAGCGGCGGAGG 866  
QY 623 -----SGRGLGGGAGAAAA-----AAAAAGAGCGGCG 652  
Db 867 GCGGLGSGGAGRGGLGGGAGAAAAAAGAGAGCGGGLGGGAGCGGAGAAAAAAGAGCGG 926  
QY 653 GGLGSGTSGPGGYPG 669  
Db 927 GGLGSGG-AGRGCGGAG 942  
RESULT 6  
AAU11798  
ID AAU11798 standard; Protein; 1626 AA.  
AC AAU11798;  
XX  
XX  
DT 26-MAR-2002 (first entry)  
DE Dragline protein 1 analogue DP-1B/his tag 16mer.  
XX  
XX Silk-like protein; SLP; transgenic plant; promoter; 5' terminator;  
KW fabric production; material construction; rope; surgical suture;  
KW flexible tie down; electrical component; implantation;  
KW Dragline protein 1; DP-1B 16mer; His tag.  
XX  
OS Nephila clavipes.  
OS Synthetic.  
XX  
PN WO200190389-A2.  
XX  
XX 29-NOV-2001.  
XX  
XX 24-MAY-2001; 2001WO-US16937.  
XX  
XX 25-MAY-2000; 2000US-206968P.  
XX  
XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
XX  
XX Yang JG;  
XX  
XX WPI; 2002-106209/14.  
XX  
XX Producing silk-like proteins in a green plant, useful in fabrics and  
PT material construction, comprises providing a silk-like protein  
PT expression cassette to a green plant -  
XX  
XX Example 1; Page 87-92; 93pp; English.  
PS  
CC The invention relates to a method of producing silk-like proteins (SLP)  
CC in a green plant by providing a green plant containing an SLP expression  
CC cassette. The green plant contains the expression cassette P-SLP-T where

P is a promoter for driving the expression of an SLP transgene and T is a 5' terminator. The method is useful for producing silks and silk-like proteins in green plants. The silks and silk-like proteins may be used in fabrics or in material construction, such as rope, surgical sutures, flexible tie downs for certain electrical components, or as a biomaterial for implantation. The method allows for more cost effective production of silk not obtained from natural or microbial sources. The present sequence is the silk protein Dragline protein (DP) 1 synthetic variant DP1B/his tag 16mer (i.e. 16 copies of DP1B with a His tag at the C-terminus) which is used as an SLP construct in the method of the invention.  
XX  
SQ Sequence 1626 AA;  
Query Match 65.5%; Score 2321.5; DB 23; Length 1626;  
Best Local Similarity 57.6%; Pred. No. 2 5e-158;  
Matches 556; Conservative 14; Mismatches 51; Indels 345; Gaps 36;  
QY 11 GRGSMASGRGLGGGAGAAAA-----AAAAAGAGCGGCGGGL 49  
Db 721 GLGSGAGRGGLGGGAGAAAAAAGAGCGGGLGSGGAGAGAAAAAAGAGCGGCGGGL 780  
QY 50 GSGTSGRGLGGGAGAAAAAAGAGAGCGGCGGGLGSGG-----SGRG 97  
Db 781 GSG-----AGRGCGGAG-----AAAAAGAGCGGCGGGLGSGGAGCGG 832  
QY 98 LGGGAGAAAA-----AAAAAGAGCGGCGGGLGSGG-----SGRGGL 136  
Db 833 LGGGAGAAAAAAGAGAGCGGGLGSGGAGAGAAAAAAGAGCGGCGGGLGSGG-----AGR 888  
QY 137 GCGGAGAAAAAAGAGAGCGGCGGGLGSGGTSRGGYGP-GQOTSGRGLGGGAGAAAA 195  
Db 889 GCGGAG-----AAAAAGAGCGGCGGGLGSGG-AGCGGCGGGLGSGGAGCGGAGAAAA 943  
QY 196 A-----AAAAAGAGCGGCGGGLGSGGTSRGGGLGGGAGAAAA 234  
Db 944 AAAGAGCGGGLGSGGAGCGGAGAAAAAAGAGCGGCGGGLGSGG-----AGRGCGGAG-----995  
QY 235 AAAAAGAGCGGCGGGLGSGG-----SGRGLGGGAGAAAA-----272  
Db 996 AAAAAGAGCGGCGGGLGSGGAGCGGCGGGLGSGGAGRGGLGGGAGAAAAAAGAGCGG 1055  
QY 273 -----AAAAAGAGCGGCGGGLGSGGTSRGGGLGGGAGAAAAAAGAGCGG 321  
Db 1056 LGSQAGCGAGAAAAAAGAGCGGCGGGLGSGG-----AGRGCGGAG-----AAAAAGAGCGG 1107  
QY 322 GCGGGLGSGTSGPGGYP-GQOTSGRGLGGGAGAAAA-----360  
Db 1108 GCGGGLGSGG-AGCGGCGGGLGSGGAGRGGLGGGAGAAAAAAGAGCGGGLGSGGAGCGA 1166  
QY 361 -AAAAAGAGCGGCGGGLGSGGTSRGGGLGGGAGAAAAAAGAGCGGCGGGLGSGG 419  
Db 1167 GAAAAAGAGCGGCGGGLGSGG-----AGRGCGGAG-----AAAAAGAGCGGCGGGLGSGG 1218  
QY 420 T-----SGRGLGGGAGAAAA-----AAAAAGAGCGGAG 446  
Db 1219 AGCGGCGGGLGSGGAGRGGLGGGAGAAAAAAGAGCGGGLGSGGAGAGAAAAAAGAGGAG 1278  
QY 447 GCGGGLGSGGTSRGGGLGGGAGAAAAAAGAGCGGCGGGLGSGGTSRGGYGP-G 505  
Db 1279 GCGGGLGSGG-----AGRGCGGAG-----AAAAAGAGCGGCGGGLGSGG-AGCGGCGG 1329  
QY 506 QOTSGRGLGGGAGAAAA-----AAAAAGAGCGGCGGGLGSGG 544  
Db 1330 SQAGRGGLGGGAGAAAAAAGAGAGCGGGLGSGGAGAGAAAAAAGAGCGGCGGGLGSGG 1389  
QY 545 GTSRGLGGCGGAGAAAAAAGAGAGCGGCGGGLGSGG-----SGRGLGG 592  
Db 1390 G-----AGRGCGGAG-----AAAAAGAGCGGCGGGLGSGGAGCGGCGGAGRGGLG 1441  
QY 593 QGAGAAAA-----AAAAAGAGCGGCGGGLGSGGAGCGGCGGCGGAGCGGAGCGA 601  
Db 1442 QGAGAAAAAAGAGAGCGGGLGSGGAGCGGCGGCGGCGGCGGAGCGGAGCGGAGCGA 1501

QY 602 AAAAAGGAGGGYGGGLGSGGT-----SRRGGLGGGAGAAA----- 638  
 DB 1502 AAAAAGGAGGGYGGGLGSGGTGGGAGGGYGGGLGSGGTGGGAGAAAAGAGAGGGGL 1561  
 QY 639 -----AAAAAGGAGGGYGGGLGSGGTGGGAGGGYGGGLGSGGTGGGAGAAAAGAGAGGGGL 685  
 DB 1562 GSGGAGGAGAAAAGGAGGGYGGGLGSGGTAGRGAGGAGAAAAGGAGGGYGGGLG 1620  
 QY 686 HHHHHH 691  
 DB 1621 HHHHHH 1626  
 RESULT 7  
 AAU11794  
 ID AAU11794 standard; protein; 1617 AA.  
 XX AC  
 XX AAU11794;  
 XX DT  
 XX 26-MAR-2002 (first entry)  
 DE Dragline protein 1 analogue DP-1B 16mer.  
 XX Orb-weaving spider; silk-like protein; SLP; transgenic plant;  
 KW promoter; 5' terminator; fabric production; material construction; rope;  
 KW surgical suture; flexible tie down; electrical component; implantation;  
 KW Dragline protein 1; DP-1B 16mer.  
 XX Nephila clavipes.  
 OS Synthetic.  
 XX WO200190389-A2.  
 XX 29-NOV-2001.  
 XX 24-MAY-2001; 2001WO-US16937.  
 XX 25-MAY-2000; 2000US-206968P.  
 XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX Yang JG;  
 XX WPI; 2002-106209/14.  
 XX Producing silk-like proteins in a green plant, useful in fabrics and  
 PT material construction, comprises providing a silk-like protein  
 PT expression cassette to a green plant -  
 XX Disclosure; Page 77-81; 93pp; English.  
 XX The invention relates to a method of producing silk-like proteins (SLP)  
 CC in a green plant by providing a green plant containing an SLP expression  
 CC cassette. The green plant contains the expression cassette P-SLP-T where  
 CC P is a promoter for driving the expression of an SLP transgene and T is a  
 CC 5' terminator. The method is useful for producing silks and silk-like  
 CC proteins in green plants. The silks and silk-like proteins may be used in  
 CC fabrics or in material construction, such as rope, surgical sutures,  
 CC flexible tie downs for certain electrical components, or as a  
 CC biomaterial for implantation. The method allows for more cost effective  
 CC production of silk not obtained from natural or microbial sources.  
 CC The present sequence is the silk protein Dragline protein (DP) 1  
 CC synthetic variant DP1B 16mer (i.e. 16 copies of DP1B) which is used as  
 CC an SLP construct in the method of the invention.  
 XX  
 SQ Sequence 1617 AA;  
 Query Match 64.9%; Score 2298; DB 23; Length 1617;  
 Best Local Similarity 61.5%; Pred. NO. 1.2e-156;  
 Matches 544; Conservative 13; Mismatches 39; Indels 288; Gaps 36;  
 QY 11 GRGSMASRRGGLGGGAGAAA-----AAAAAGGAGGGYGGGL 49

DB 12 GLGSGAGRGLGGGAGAAAAGAGAGGGYGGGLGSGGTGGGAGAAAAGAGAGGGYGGGL 71  
 QY 50 GSQGTSGRGLGGGAGAAAAGAGAGGGYGGGLGSGGT-----SRRGG 97  
 DB 72 GSQGT-----AGRGGGAG-----AAAAAGGAGGGYGGGLGSGGTGGGAGRG 123  
 QY 98 LGGGAGAAA-----AAAAAGGAGGGYGGGLGSGGTSGRGL 136  
 DB 124 LGGGAGAAAAGAGAGGGYGGGLGSGGTGGGAGAAAAGAGAGGGYGGGLGSG 179  
 QY 137 GGAGAGAAAAGAGAGGGYGGGLGSGGTSGPGYGP-GQOTSGRGLGGGAGAAA 195  
 DB 180 GGAG-----AAAAAGGAGGGYGGGLGSGGT-AGQGYGGGLGSGGTGGGAGAAA 234  
 QY 196 A-----AAAAAGGAGGGYGGGLGSGGTSGRGLGGGAGAAA 234  
 DB 235 AAAAGGAGGGYGGGLGSGGT-----SRRGGLGGGAGAAA----- 286  
 QY 235 AAAAGGAGGGYGGGLGSGGT-----SRRGGLGGGAGAAA----- 272  
 DB 287 AAAAGGAGGGYGGGLGSGGTGGGAGAAAAGAGAGGGYGGGLGSGGTGGGAGAAAAGAGAG 346  
 QY 273 -----AAAAAGGAGGGYGGGLGSGGTSGRGLGGGAGAAA----- 360  
 DB 347 LGSGAGAGAGAGAGAGAGGGYGGGLGSGGT-----AGRGGGAG-----AAAAAGGAG 398  
 QY 322 GYGGLGSGGTSGPGYGP-GQOTSGRGLGGGAGAAA----- 360  
 DB 399 GYGGLGSGGT-AGQGYGGGLGSGGTGGGAGAAAAGAGAGGGYGGGLGSGGTGGGAG 457  
 QY 361 -AAAAAGGAGGGYGGGLGSGGTSGRGLGGGAGAAAAGAGAGGGYGGGLGSG 419  
 DB 458 GAAAAAGGAGGGYGGGLGSGGT-----AGRGGGAG-----AAAAAGGAGGGYGGGLGSG 509  
 QY 420 T-----SRRGGLGGGAGAAA-----AAAAAGGAG 446  
 DB 510 AGGGYGGGLGSGGTGGGAGAAAAGAGAGGGYGGGLGSGGTGGGAGAAAAGAGAG 569  
 QY 447 QGYGGGLGSGGTSGRGLGGGAGAAAAGAGAGGGYGGGLGSGGTSGPGYGP-G 505  
 DB 570 QGYGGGLGSGGT-----AGRGGGAG-----AAAAAGGAGGGYGGGLGSGGT-AGQGYGGGLG 620  
 QY 506 QGTSGRGLGGGAGAAA-----AAAAAGGAGGGYGGGLGSG 544  
 DB 621 SQAGRGLGGGAGAAAAGAGAGGGYGGGLGSGGTGGGAGAAAAGAGAGGGYGGGLGSG 680  
 QY 545 GTSRGLGGGAGAAAAGAGAGGGYGGGLGSGGT-----SRRGGLGG 592  
 DB 681 G-----AGRGGGAG-----AAAAAGGAGGGYGGGLGSGGTGGGAGRGGLGG 732  
 QY 593 QGAGAAA-----AAAAAGGAGGGYGGGLGSGGTSGRGLGG 631  
 DB 733 QGAGAAAAGAGAGGGYGGGLGSGGTGGGAGAAAAGAGAGGGYGGGLGSGGT-----AGRG 788  
 QY 632 GAGAAAAGAGAGGGYGGGLGSGGTSGPGYGP-GQOTSG 674  
 DB 789 GAG-----AAAAAGGAGGGYGGGLGSGGT-AGQGYGGGLGSGGT 827  
 RESULT 8  
 AAU11797  
 ID AAU11797 standard; Protein; 818 AA.  
 XX AC  
 XX AAU11797;  
 XX DT  
 XX 26-MAR-2002 (first entry)  
 DE Dragline protein 1 analogue DP-1B/his tag 8mer.  
 XX Silk-like protein; SLP; transgenic plant; promoter; 5' terminator;  
 KW fabric production; material construction; rope; surgical suture;  
 KW flexible tie down; electrical component; implantation;





QY	6	GGQOMGRGSMASGRGGLGGGACAGAAAAAAGGAGAGGAGGGYGGGLGSGQTSGRGGLGGQGA	65
Db	9	GGGAGGGG-----GYGLGGGAGGAGAAAAAAGGAGGGYGGGLGSGQ-----AGRGGQGA	60
QY	66	GAIAAAAAAAGAGAGGGYGGGLGSGQTSGRGGLGGGAGAAAAAAGGAGAGGGYGGGL	125
Db	61	G-----AAAAAGAGAGGGYGGGLGSGQ-AGRGGLGGGAG-----AAAAAGAGAGGGYGGGL	111
QY	126	GSQGTSGRGLGGGAGAAAAAAGGAGAGGGYGGGLGSGQTSGRGGLGGGAGGGYGGGL	185
Db	112	GSQG-----AGRGGQAG-----AAAAAGGAGGGYGGGLGSGQ-----AGRGG	150
QY	186	LGSGGAGAAAAA-----AGGAGGGYGGGLGSGQTSGRGGLGGGAGAAAAA	236
Db	151	LGSGGAGAAAAAGGAGGGYGGGLGGGAGGGYGGGLGSGQ-AGRGGLGGGAG-----AA	205
QY	237	AAAGGAGGGYGGGLGSGQTSGRGGLGGGAGAAAAAAGGAGGGYGGGLGSGQTS	296
Db	206	AAAGGAGG-----GGLGGGAGGAGAAAAAAGGAGGGYGGGLGSGQ-----	248
QY	297	RGGLGGGAGAAAAAAGGAGAGGGYGGGLGSGQTSGRGGLGGGAGGGYGGGLGSGQ	355
Db	249	AGRGGQAG-----AAAAAGGAGGGYGGGLGSGQ-AGGGYGGGLGSGQAGRGGLGGQGA	302
QY	356	GAIAA-----AAAAAGAGGGYGGGLGSGQTSGRGGLGGGAG	395
Db	303	GAIAAAGGAGGGGAGGGGAGGAGAAAAAAGGAGGGYGGGLGSGQ-AGRGGLGGGAG	360
QY	396	AAAAAAGGAGGGYGGGLGSGQTSGRGGLGGGAGAAAAAAGGAGGGYGGGLGSG	455
Db	361	-----AAAAAGGAGGGYGGGLGSGQ-----AGRGGQAG-----AAAAAGGAGGGYGGGLGSG	409
QY	456	QGTSGRGLGGGAGAAAAAAGGAGGGYGGGLGSGQTSGRGGLGGGAGGGYGGGLGSG	515
Db	410	QG-AGRGGLGGGAG-----AAAAAGGAGGGYGGGLGSGQ-----AGRG	448
QY	516	GOGAGAAAAAAGGAGGGYGGGLGSGQTSGRGGLGGGAGGGYGGGLGSG-----	567
Db	449	GOGAG-----AAAAAGGAGGGYGGGLGSGQ-----AGRGGQAGAAAAAAGGAGGGYGGGL	500
QY	568	-AGGAGGGYGGGLGSGQTSGRGGLGGGAGAAAA-----AAAAA	606
Db	501	GOGAGGGYGGGLGSGQ-AGRGGLGGGAGAAAAAAGGAGGGYGGGLGSGAGGAGAAAAA	559
QY	607	GAGGGYGGGLGSGQTSGRGGLGGGAGAAAAAAGGAGGGYGGGLGSGQTS---SGPG	664
Db	560	GAGGGYGGGLGSGQ-----AGRGGQAG-----AAAAAGGAGGGYGGGLGSGAGRGGLG	611
QY	665	GYPG 669	
Db	612	GOGAG 616	
RESULT 11			
AAW50047			
ID	AAW50047 standard; Protein; 676 AA.		
XX	AAW50047;		
AC			
XX			
DT	18-SEP-2002 (first entry)		
XX			
DE	N. clavipes spidroin synthetic homologue SOL protein #2.		
XX			
KW	Spidroin; spider; silk; fibre; film; membrane; wound; filter; SOL.		
XX			
OS	Synthetic.		
XX			
FH	Key		
FT	1..28 Location/Qualifiers		
FT	Peptide		
FT	29..659 /label= LeB4_signal_peptide		
FT	Protein		
FT	/note= "Synthetic spidroin homologue SOL"		
FT	Region		
FT	660..676		

FT	Domain	/note= "c-myc-tag" 673..676 /note= "ER retention signal"
FT		
FT		
XX	DE10113781-A1.	
XX	13-DEC-2001.	
PD		
XX	21-MAR-2001; 2001DE-1013781.	
PF		
XX	09-JUN-2000; 2000DE-1028212.	
PR		
PR	24-OCT-2000; 2000DE-1053478.	
XX		
PA	(IPKP-) IPK INST PFLANZENGENETIK & KULTURPFLANZE.	
XX		
PI	Scheller J, Conrad U, Grosse F, Guehrs K;	
XX	WPI; 2002-123561/17.	
DR		
XX		
PT	New DNA encoding synthetic spider silk protein, useful e.g. for closing wounds, comprises modules that encode repeating units of spidroin proteins	
PT		
XX		
XX		
PS	Example 1; Fig 10A; 88pp; German.	
XX		
CC	This invention describes a novel DNA sequence, encoding a synthetic spider silk protein, comprising modules, each comprising a group of sequentially arranged oligonucleotides, each oligonucleotide encoding a repeating unit of a spidroin protein. The synthetic protein has at least 84% homology with the Nephila clavipes spidroin protein and is used to produce synthetic fibres, films and/or membranes, particularly: (i) for medical use, especially to close wounds and/or to support or cover artificial organs; (ii) as adhesion surfaces for culturing cells; and (iii) as filters. The synthetic proteins are very similar to native spider silk proteins; can be prepared on a large scale and can be spun to fibres with excellent mechanical properties (strength and elasticity). CC	
CC	They retain water solubility after long-term boiling in aqueous solutions and since they are also soluble in organic solvents but precipitated at high salt concentration, they are easily extracted and purified. The modular construction of the invention facilitates incorporation of additional peptide-encoding sequences, e.g. to simplify purification or modulate solubility. This sequence represents a construct composed of the LeB4 signal peptide, N. clavipes spidroin-1 synthetic homologue SOL, a c-Myc-tag and an endoplasmic reticulum (ER)-retention signal described in the invention.	
XX		
SQ	Sequence 676 AA;	

Query Match		60.7%; Score 2151; DB 23; Length 676;
Best Local Similarity		68.6%; Pred. No. 1.9e-146;
Matches 497; Conservative 9; Mismatches 41; Indels 178; Gaps 32;		
QY	6	GGQOMGRGSMASGRGGLGGGAGAAAAAAGGAGAGGGYGGGLGSGQTSGRGGLGGQGA 65
Db	37	GGGAGGG-----GYGLGGGAGGAGAAAAAAGGAGGGYGGGLGSGQ-----AGRGGQGA 88
QY	66	GAIAAAAAAAGAGAGGGYGGGLGSGQTSGRGGLGGGAGAAAAAAGGAGAGGGYGGGL 125
Db	89	G-----AAAAAGAGAGGGYGGGLGSGQ-AGRGGLGGGAG-----AAAAAGAGAGGGYGGGL 139
QY	126	GSQGTSGRGLGGGAGAAAAAAGGAGAGGGYGGGLGSGQTSGRGGLGGGAGGGYGGGL 185
Db	140	GSQG-----AGRGGQAG-----AAAAAGGAGGGYGGGLGSGQ-----AGRGG 178
QY	186	LGSGGAGAAAAA-----AGGAGGGYGGGLGSGQTSGRGGLGGGAGAAAAA 236
Db	179	LGSGGAGAAAAAGGAGGGYGGGLGGGAGGGYGGGLGSGQ-AGRGGLGGGAG-----AA 233
QY	237	AAAGGAGGGYGGGLGSGQTSGRGGLGGGAGAAAAAAGGAGGGYGGGLGSGQTS 296
Db	234	AAAGGAGG-----GGLGGGAGGAGAAAAAAGGAGGGYGGGLGSGQ--- 276
QY	297	RGGLGGGAGAAAAAAGGAGGGYGGGLGSGQTSGRGGLGGGAGGGYGGGLGSGQ 355









```

      ||| | | :|| ||||| ||||| | ||||| ||||| ||||| |||||
Db 476 GYGGLGSGAGR---GGGAGAAAAAYGAGQEGIRGQAGQGGYGLGSGQ-SGRGGL 531
QY 553 GGQAGAAAA-----AAAAAGGAGQGGYGLGSGQTSGRGGLGG 592
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 532 GGQAGAAAAAGGAGQGGGLGGQAGQAGAAAAAGGVRRQGGYGLGSGQ----AGRG 587
QY 593 QGAGAAAAAGGAGQGGYGLGSGQTSGRGGLGGQAGAAAAAGGAGQGGY 652
      ||| | ||||| ||||| || ||||| ||||| ||||| |||||
Db 588 QGAG----AAAAAGGAGQGGYGLGGQGV-CRGGLGGQAG-----AAAAGGAGQGGY 636
QY 653 GGLGS 657
      ||:|
Db 637 GGVS 641
```

Search completed: December 18, 2002, 16:17:19  
Job time : 53.1374 secs





QY 653 GGLGSGTSGPGYGGQQTSGIRAPSTS 683  
Db 637 GGVGSGASAA-----SAAASRLSPQAS 659

## RESULT 3

US-09-034-177-3  
; Sequence 3, Application US/09034177  
; Patent No. 6127146  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: HUMAN FIBROUS PROTEIN  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; APPLICATION DATA:  
; APPLICATION NUMBER: US/09/034,177  
; FILING DATE: HEREWITH  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0486 US  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 747 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: GI 1174414  
; US-09-034-177-3

Query Match 60.3%; Score 2138; DB 3; Length 747;  
Best Local Similarity 56.3%; Pred. No. 3.9e-147;  
Matches 498; Conservative 20; Mismatches 49; Indels 184; Gaps 35;  
QY 25 QGAGAAAAAAGAGGAGGAGGGLGSGQT--SGRGLGGGAG-AAAAAAGAGGAG 81  
Db 1 QGAG-----AAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 56  
QY 82 GGVGGLGSGTSGRGLGGGAGAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 141  
Db 57 GGVGGLGSGQ-----AGRGCGAG-----AAAAAGGAGGAGGAGGAGGAG 107  
QY 142 GAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 201  
Db 108 G-----AAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 142  
QY 202 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 252  
Db 143 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 201

QY 253 SQGTSRGLGGGAGAAAAA-----AAAAAGAGGAGGAGGAGGAG 291  
Db 202 SQG-AGRGLGGGAGAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 260  
QY 292 QGTSRGLGGGAGAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 350  
Db 261 QG-----AGRGEGAG-----AAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 311  
QY 351 GGGAGAAAA-----AAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 392  
Db 312 GGGAGAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 370  
QY 393 AGAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 452  
Db 371 AG-----AVAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 419  
QY 453 LGSQTSRGLGGGAGAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 499  
Db 420 LGNG-AGRGLGGAG-----AAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 475  
QY 500 GGYGP-QQTSRGLGGGAGAAAAA-----AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 552  
Db 476 GGVGGLGSGAGR---GGGAGAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 531  
QY 553 GGGAGAAAA-----AAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 592  
Db 532 GGGAGAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 587  
QY 593 QGAGAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 652  
Db 588 QGAG-----AAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 636  
QY 653 GGLGSGTSGPGYGGQQTSGIRAPSTS 683  
Db 637 GGVGSGASAA-----SAAASRLSPQAS 659

## RESULT 4

US-08-556-978B-19  
; Sequence 19, Application US/08556978B  
; Patent No. 6268169  
; GENERAL INFORMATION:  
; APPLICANT: FAHNESTOCK, STEPHEN F.  
; TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED  
; TITLE OF INVENTION: SPIDER SILK ANALOGS  
; NUMBER OF SEQUENCES: 107  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
; STREET: 1007 MARKET STREET  
; CITY: WILMINGTON  
; STATE: DELAWARE  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 19898  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.50 INCH  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: MICROSOFT WINDOWS 95  
; SOFTWARE: MICROSOFT WORD FOR WINDOWS 95  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/556, 978B  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/077,600  
; FILING DATE: JUNE 15, 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FLOYD, LINDA AXAMETHY  
; REGISTRATION NUMBER: 33,692  
; REFERENCE/DOCKET NUMBER: CR-9389-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 302-892-8112  
; TELEFAX: 302-773-0164

; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 651 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-556-978B-19

Query Match 60.3%; Score 2135; DB 4; Length 651;  
Best Local Similarity 68.3%; Pred. No. 5.7e-147;  
Matches 495; Conservative 15; Mismatches 39; Indels 176; Gaps 34;

QY 25 QGAGAAAAAAGAGAGAGGGTGGTGGT--SRRGGLGGGAG-AAAAAAGAGAG 81  
Db 1 QGAG-----AAAAAGAGAGGGTGGTGGGAGGGTGGGAGAGAGAGAGAGAG 56  
QY 82 GYGGLGGTSGRGLGGGAGAGAAAAAAGAGAGGGTGGTGGTSGRGLGGGAG 141  
Db 57 GYGGLGGTSGRGLGGGAGAG-----AAGAGAGAGAGAGAGAGAGAGAG 107  
QY 142 GAAAAAAGAGAGAGAGGGTGGTGGTSGRGLGGGAGAGAGAGAGAGAGAG 201  
Db 108 G-----AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 142  
QY 202 AGAGAGGGTGGTGGTSGRGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 252  
Db 143 AGAGAGGGTGGTGGTSGRGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 107  
QY 253 SGTSGRGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 291  
Db 202 SGTSGRGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 260  
QY 292 QGTSGRGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 350  
Db 261 QG-----AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 311  
QY 351 GGGAG 392  
Db 312 GGGAG 370  
QY 393 AG 452  
Db 371 AG-----AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 419  
QY 453 LGSQGTSGRGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 499  
Db 420 LGNQG-AGRGLGGGAG-----AAGAGAGAGAGAGAGAGAGAGAGAGAG 475  
QY 500 GYGTP-GQTSRGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 552  
Db 476 GYGTP-GQTSRGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 531  
QY 553 GGGAG 592  
Db 532 GGGAG 587  
QY 593 QGAG 652  
Db 588 QGAG-----AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 636  
QY 653 GGLGS 657  
Db 637 GGVGS 641

RESULT 5  
US-09-247-806-1  
; Sequence 1, Application US/09247806  
; Patent No. 6280747  
; GENERAL INFORMATION:  
; APPLICANT: PHILLIPPE, Michel  
; APPLICANT: GARSON, Jean-Claude

; APPLICANT: ARRAUDEAU, Jean-Pierre  
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT  
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN  
; FILE REFERENCE: 6388-0365-0  
; CURRENT APPLICATION NUMBER: US/09/247,806  
; CURRENT FILING DATE: 1999-02-11  
; EARLIER APPLICATION NUMBER: FR 98/01614  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 651  
; TYPE: PRT  
; ORGANISM: Nephila clavipes  
US-09-247-806-1

Query Match 60.3%; Score 2135; DB 4; Length 651;  
Best Local Similarity 68.3%; Pred. No. 5.7e-147;  
Matches 495; Conservative 15; Mismatches 39; Indels 176; Gaps 34;

QY 25 QGAGAAAAAAGAGAGAGGGTGGTGGT--SRRGGLGGGAG-AAAAAAGAGAG 81  
Db 1 QGAG-----AAAAAGAGAGGGTGGTGGGAGGGTGGGAGAGAGAGAGAGAG 56  
QY 82 GYGGLGGTSGRGLGGGAGAGAAAAAAGAGAGGGTGGTGGTSGRGLGGGAG 141  
Db 57 GYGGLGGTSGRGLGGGAGAG-----AAGAGAGAGAGAGAGAGAGAGAG 107  
QY 142 GAAAAAAGAGAGAGAGGGTGGTGGTSGRGLGGGAGAGAGAGAGAGAGAG 201  
Db 108 G-----AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 142  
QY 202 AGAGAGGGTGGTGGTSGRGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 252  
Db 143 AGAGAGGGTGGTGGTSGRGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 201  
QY 253 SGTSGRGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 291  
Db 202 SGTSGRGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 260  
QY 292 QGTSGRGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 350  
Db 261 QG-----AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 311  
QY 351 GGGAG 392  
Db 312 GGGAG 370  
QY 393 AG 452  
Db 371 AG-----AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 419  
QY 453 LGSQGTSGRGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 499  
Db 420 LGNQG-AGRGLGGGAG-----AAGAGAGAGAGAGAGAGAGAGAGAGAG 475  
QY 500 GYGTP-GQTSRGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 552  
Db 476 GYGTP-GQTSRGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 531  
QY 553 GGGAG 592  
Db 532 GGGAG 587  
QY 593 QGAG 652  
Db 588 QGAG-----AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 636  
QY 653 GGLGS 657  
Db 637 GGVGS 641













```

RESULT 14
US-08-556-978B-61
; Sequence 61, Application US/08556978B
; Patent No. 6268169
; GENERAL INFORMATION:
; APPLICANT: FAHNESTOCK, STEPHEN F.
; TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED
; TITLE OF INVENTION: SPIDER SILK ANALOGS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD FOR WINDOWS 95
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/556,978B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/077,600
; FILING DATE: JUNE 15, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA AXAMETHY
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CR-9389-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 714 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-556-978B-61

Query Match 38.0%; Score 1346.5; DB 4; Length 714;
Best Local Similarity 50.0%; Pred. No. 3.4e-90;
Matches 367; Conservative 26; Mismatches 248; Indels 93; Gaps 35;

QY 2 ASMTGGQMGKSGMASGRGLGGQAG----GAGGGYGLGSGQTSRGLG-GQGAGAAAAAA 111
Db 10 AAAGPGQGGPGGYGPGQGGPGGYGPGQGGPGGSAAAAAAAGPGGYGPGQGGPGGSA 69
QY 57 RGLGGQAGAAAAAAG----GAGGGYGLGSGQTSRGLG-GQGAGAAAAAA 111
Db 70 PQQGPSPGSAAAAAAAGPGGYGPGQGGPGGYGPGQGGPGGSA 128
QY 112 AAAGAGGGYGLGSGQTSRGLG-GQGAGAAAAAAGAGAGGGYGLGSGQTS 170
Db 129 AAAGPGQGGPGGYGPGQGGPGGYGPGQGGPGGSAAAAAAAGPGGYG 183
QY 171 PGYGPQQTSGRGLGGQAGAAAAAAG----GAGGGYGLGSGQTSRGLG- 225
Db 184 PGYGPQQTSGRGLGGQAGAAAAAAGPGGYGPGQGGPGGYGPGQGGPGGYG 235
QY 226 GQAGAAAAAAGAGAGGGYGLGSGQTSRGLG-GQGAGAAAAAAGAGAGG 284
Db 236 GQGGPGGSAAAAAAAGPGQGGPGGYGPGQGGPGGYGPGQGGPGGSA 294
QY 285 YGGLGSGQTSRGLGGQAGAAAAAAGAGAGGGYGLGSGQTSRGLGPGQQT 344
Db 295 YGYP-GQGGPGGYGPGQGGPGGSAAAAAAAGPGGYG----PGQGGPGGYGPGQGG 348

RESULT 15
US-09-247-806-10
; Sequence 10, Application US/09247806
; Patent No. 6280747
; GENERAL INFORMATION:
; APPLICANT: PHILLIPPE, Michel
; APPLICANT: GARSON, Jean-Claude
; APPLICANT: ARRAUDEAU, Jean-Pierre
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
; TITLE OF INVENTION: SPIDER SILK OR AN
; TITLE OF INVENTION: ANALOG
; FILE REFERENCE: 6388-0365-0
; CURRENT APPLICATION NUMBER: US/09/247,806
; CURRENT FILING DATE: 1999-02-11
; EARLIER APPLICATION NUMBER: FR 98/01614
; EARLIER FILING DATE: 1998-02-11
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 714
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:protein
US-09-247-806-10

Query Match 38.0%; Score 1346.5; DB 4; Length 714;
Best Local Similarity 50.0%; Pred. No. 3.4e-90;
Matches 367; Conservative 26; Mismatches 248; Indels 93; Gaps 35;

QY 2 ASMTGGQMGKSGMASGRGLGGQAG----GAGGGYGLGSGQTSRGLG-GQGAGAAAAAA 403
Db 349 PGYGPQGGPGGPGSAAAAAA-GPQGGPGGYG-PGQGGPGGYGPGQGGPGGSA 406
QY 404 AGGAGGGYGLGSGQTSRGLGGQAGAAAAAAGAGAGGGYGLGSGQTSRGLG- 452
Db 407 AAAGPGGYGPGQGGPGGYGPGQGGPGGSAAAAAAAGPGGYGPGQGGPGGYG 463
QY 463 ---GLGGQAG-----AAAAAAGAGAGGGYG-----GLGSGQTS----- 497
Db 464 GQGGPGGYGPGQGGPGGSAAAAAAAGPGQGGPGGYGPGQGGPGGPGGSA 523
QY 498 -----GPGGYGPGQGGPGGAGAAAAAAGAGAGGGYGLGSGQTSRGLG- 550
Db 524 AAAAAAGPGGYGPGQGGPGGYGPGQGGPGGSAAAAAAAGPGGYGPGQGGPGGYG 582
QY 551 ---GLGGQAG-----AAAAAAGAGAGGGYGLGSGQTSRGLG-GQGAGAAAAAA 602
Db 583 GQGGPGGYGPGQGGPGGSAAAAAAAGPGQGGPGGYG-PGQGGPGGYGPGQGGPGGSA 641
QY 603 AAGAGAGGGYG-GLGSGQTSRGLGGQAGAAAAAAG-----GAGGGYGLGSG 657
Db 642 AAAAAAGPGGYGPGQGGPGGYGPGQGGPGGSAAAAAAAGPGGYGPGQGGPGGYG- 700
QY 658 QGTSGPGGYGPGQ 671
Db 701 PGQGGPGGYGPGQ 714

US-09-247-806-10
; Sequence 10, Application US/09247806
; Patent No. 6280747
; GENERAL INFORMATION:
; APPLICANT: PHILLIPPE, Michel
; APPLICANT: GARSON, Jean-Claude
; APPLICANT: ARRAUDEAU, Jean-Pierre
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
; TITLE OF INVENTION: SPIDER SILK OR AN
; TITLE OF INVENTION: ANALOG
; FILE REFERENCE: 6388-0365-0
; CURRENT APPLICATION NUMBER: US/09/247,806
; CURRENT FILING DATE: 1999-02-11
; EARLIER APPLICATION NUMBER: FR 98/01614
; EARLIER FILING DATE: 1998-02-11
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 714
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:protein
US-09-247-806-10

Query Match 38.0%; Score 1346.5; DB 4; Length 714;
Best Local Similarity 50.0%; Pred. No. 3.4e-90;
Matches 367; Conservative 26; Mismatches 248; Indels 93; Gaps 35;

QY 2 ASMTGGQMGKSGMASGRGLGGQAG----GAGGGYGLGSGQTSRGLG-GQGAGAAAAAA 56
Db 10 AAAGPGQGGPGGYGPGQGGPGGYGPGQGGPGGSAAAAAAAGPGGYGPGQGGPGGSA 69
QY 57 RGLGGQAGAAAAAAG----GAGGGYGLGSGQTSRGLG-GQGAGAAAAAA 111
Db 70 PQQGPSPGSAAAAAAAGPGGYGPGQGGPGGYGPGQGGPGGSA 128
QY 112 AAAGAGGGYGLGSGQTSRGLG-GQGAGAAAAAAGAGAGGGYGLGSGQTS 170
Db 129 AAAGPGQGGPGGYGPGQGGPGGYGPGQGGPGGSAAAAAAAGPGGYG 183
QY 171 PGYGPQQTSGRGLGGQAGAAAAAAG----GAGGGYGLGSGQTSRGLG- 225
Db 184 PGYGPQQTSGRGLGGQAGAAAAAAGPGGYGPGQGGPGGYGPGQGGPGGYG 235
QY 226 GQAGAAAAAAGAGAGGGYGLGSGQTSRGLG-GQGAGAAAAAAGAGAGG 284
Db 236 GQGGPGGSAAAAAAAGPGQGGPGGYGPGQGGPGGYGPGQGGPGGSA 294
QY 285 YGGLGSGQTSRGLGGQAGAAAAAAGAGAGGGYGLGSGQTSRGLGPGQQT 344
Db 295 YGYP-GQGGPGGYGPGQGGPGGSAAAAAAAGPGGYG----PGQGGPGGYGPGQGG 348
```

Db 184 PGYGPQQ-----GPSGSAAAAAAAAAAGPGYGPQQPGGYG-PQQQPGGYGP 235  
QY 226 GQAGAAAAAAGAGGGYGLGSGTSGRGLG-CQGAGAAAAAAGAGAGG 284  
Db 236 GQQGPGGSAAGPGQPGGYG-PQQQPGYGPQQPGPGSAAGAAAAAGP 294  
QY 285 YGGLGSGTSGRGLGCGAGAAAAAAGAGGOGGYGGLGSGTSGPGYGPQQ 344  
Db 295 GYGP-GQQPGGYGP-CQQPGPGSAAGAAAAAGPGYG----PQQQPGYGPQQ 348  
QY 345 SGRGLGCGAGAAAAAAGAGGOGGYGGLGSGTSGRGLG-CQGAGAAAAA 403  
Db 349 PGYGPQQPGPGSAAGAAAA-GPGQPGGYG-PQQQPGYGPQQPGPGSA 406  
QY 404 AGAGGGYGLGSGTSGRGLGCGAGAAAAAAGAGGOGGYGGLGSGTSGR 462  
Db 407 AAAAGPGYGP-GQQPGGYGP-GQQPGPGSAAGAAAAAGPGYGP-GQQPGGYGP 463  
QY 463 ---GLGGAG-----AAAAAAGAGGOGGYG-----GLGSGTSG- 497  
Db 464 GQQPGYGPQQPGPGSAAGAAAAAGPGQPGGYGPQQPGYGPQQPGPGSA 523  
QY 498 -----GPGYGPQQTSGRGLGCGAGAAAAAAGAGGOGGYGGLGSGTSGR 550  
Db 524 AAAAAAGPGYGPQQPGGYGPQQPGPGSAAGAAAAAGPGYGP-GQQPGGYGP 582  
QY 551 ---GLGGAG-----AAAAAAGAGGOGGYGGLGSGTSGRGLG-CQGAGAAAA 602  
Db 583 GQQPGYGPQQPGPGSAAGAAAAAGPGQPGGYG-PQQQPGYGPQQPGPGSA 641  
QY 603 AAGAGGOGGYG-GLGSGTSGRGLGCGAGAAAAAAG-----GAGGGYGLG 657  
Db 642 AAAAAAGPGYGPQQPGGYGPQQPGPGSAAGAAAAAGPGYGPQQPGGYG- 700  
QY 658 QGTSGPGYGPQQ 671  
Db 701 PGQPGGYGPQQ 714

**This Page Blank (uspto)**



SUMMARIES						
Result No.	Score	Query Match	Length	DB	ID	Description
1	2135	60.3	651	10	US-09-861-597-1	Sequence 1, Appli
2	2076	58.6	606	10	US-09-861-597-4	Sequence 4, Appli
3	1972.5	55.7	606	10	US-09-861-597-8	Sequence 8, Appli
4	1961.5	55.4	606	10	US-09-861-597-6	Sequence 6, Appli
5	1346.5	38.0	714	10	US-09-861-597-10	Sequence 10, Appli
6	1098.5	31.0	529	10	US-09-861-597-2	Sequence 2, Appli
7	863.5	28.4	1894	12	US-10-052-586-97	Sequence 97, Appli
8	742	20.9	4679	10	US-09-804-898-2	Sequence 2, Appli
9	687.5	13.4	1300	12	US-10-052-586-269	Sequence 269, App
10	675.5	19.1	1497	9	US-09-060-854B-2	Sequence 2, Appli
11	672	19.0	3907	9	US-10-029-217A-24	Sequence 24, Appli
12	662	18.7	2211	9	US-10-086-861-1	Sequence 1, Appli
13	652	18.4	1400	10	US-09-879-957-37	Sequence 37, Appli
14	651.5	18.4	720	10	US-09-756-071B-20	Sequence 20, Appli
15	610	17.2	4440	12	US-10-052-586-525	Sequence 525, App
16	607	17.1	357	10	US-09-864-761-35807	Sequence 35807, A
17	607	17.1	1743	12	US-10-052-586-451	Sequence 451, App
18	588	16.6	1422	10	US-09-735-933-1	Sequence 1, Appli
19	570.5	16.1	1200	10	US-09-826-508-3	Sequence 3, Appli





```
; PRIOR APPLICATION NUMBER: 09/247,806
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: FR 98/01614
; PRIOR FILING DATE: 1998-02-11
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 714
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:protein
; US-09-861-597-10

Query Match      38.0%; Score 1346.5; DB 10; Length 714;
Best Local Similarity 50.0%; Pred. No. 1.3e-73;
Matches 367; Conservative 26; Mismatches 248; Indels 93; Gaps 35;

Qy  2 ASMTGGQMGKRSWASRGGLGGGAG-----AAAAAAGAGGAGGGY-GLGSQGTSG 56
Db  10 AAAAAAGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 69
Qy  57 RGLGGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 111
Db  70 PQQGGPSGSAASAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 128
Qy  112 AAAGCAGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 170
Db  129 AAAAAAGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 183
Qy  171 PGGYGPQGTSGRGLGGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 225
Db  184 PGGYGPQGTSGRGLGGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 235
Qy  226 GCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 284
Db  236 GCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 294
Qy  285 YGGLGSGQTSRGLGGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 344
Db  295 YGYP-GQGGPGYGP-GQGGPSGSAASAAAAAAGAGAGAGAGAGAGAGAGAGAG 348
Qy  345 SGRGLGGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 403
Db  349 PGGYGPQGTSGRGLGGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 406
Qy  404 AGGAGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 462
Db  407 AAGAGPGYGP-GQGGPGYGP-GQGGPSGSAASAAAAAAGAGAGAGAGAGAGAGAG 463
Qy  463 ---GLGGGAG-----AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 497
Db  464 GQGGPGYGPQGGPGSGSAASAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 523
Qy  498 -----GPGYGPQGTSGRGLGGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAG 550
Db  524 AAAAAAGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 582
Qy  551 ---GLGGGAG-----AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 602
Db  583 GQGGPGYGPQGGPGSGSAASAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 641
Qy  603 AAAAAAGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 657
Db  642 AAAAAAGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 700
Qy  658 GTSRGGYGPQGG 671
Db  701 PQGGPGYGPQGG 714

RESULT 6
US-09-861-597-2
```

```
; Sequence 2, Application US/09861597
; Patent No. US20020064539A1
; GENERAL INFORMATION:
; APPLICANT: PHILLIPPE, Michel
; APPLICANT: GARSON, Jean-Claude
; APPLICANT: ARRAUDEAU, Jean-Pierre
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
; TITLE OF INVENTION: ANALOG
; FILE REFERENCE: 6388-0365-0
; CURRENT APPLICATION NUMBER: US/09/861,597
; CURRENT FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: 09/247,806
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: FR 98/01614
; PRIOR FILING DATE: 1998-02-11
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Nephila clavipes
; US-09-861-597-2

Query Match      31.0%; Score 1098.5; DB 10; Length 529;
Best Local Similarity 44.4%; Pred. No. 4.9e-59;
Matches 309; Conservative 25; Mismatches 159; Indels 203; Gaps 33;

Qy  6 GQQMGKRSWASRGGLGGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 65
Db  5 GPGQGGPGYGPQGGPGSGSAASAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 52
Qy  66 CAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 123
Db  53 -----PQGGPGYGP-GQGGPGYGP-GQGGPGYGP-GQGGPGYGP-GQGGPGYGP 92
Qy  124 -----GLGSQGTSGRGLGGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAG 176
Db  93 QGGPGYGPQGGPGSGSAASAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 141
Qy  177 GQGTSGRGLG-----GCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 280
Db  142 GQ--GPGYGPQGGPGSGSAASAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 198
Qy  226 CCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 254
Db  199 GPGSAASAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 254
Qy  281 AGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 340
Db  255 QGGYGP-GQGGPGYGP-----SAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 298
Qy  341 GQGTSGRGLGGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 400
Db  299 GQ-----GPGAGSAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 326
Qy  401 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 459
Db  327 ---PQGGPGYGP-----PQGGPGYGP-GPGSAASAAAAAGAGAGAGAGAGAGAG 375
Qy  460 GRGLGGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 519
Db  376 GPG-----SASAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 414
Qy  520 CAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 577
Db  415 GSASAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 468
Qy  578 GLGSQGTSGRGLGGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 637
Db  469 P-AQGGPGYGP-----IAASASAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 507
Qy  638 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 673
```

```

Db      508 A08AGA-----GSAGYGPQSQAS 525

RESULT 7
US-10-052-586-97
; Sequence 97, Application US/10052586
; Patent No. US20020127584A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C1
; CURRENT APPLICATION NUMBER: US/10/052,586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063564
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063734
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063870
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/064103
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066120
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/066466
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066772
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/069335
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/069425
; PRIOR FILING DATE: 1997-12-12
; PRIOR APPLICATION NUMBER: 60/069870
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 60/068017
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081838
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082568
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082569
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082704
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082797
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083495
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083496
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083499
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083559
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/084366
; PRIOR FILING DATE: 1998-05-05
; PRIOR APPLICATION NUMBER: 60/084414
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084639
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084640
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084643
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085582
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085700
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/086023
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/086392
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/086486
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087098
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087208
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-05-28

```

;; PRIOR FILING DATE: 1998-06-02  
;; PRIOR APPLICATION NUMBER: 60/087759  
;; PRIOR FILING DATE: 1998-06-02  
;; PRIOR APPLICATION NUMBER: 60/087827  
;; PRIOR FILING DATE: 1998-06-03  
;; PRIOR APPLICATION NUMBER: 60/088025  
;; PRIOR FILING DATE: 1998-06-04  
;; PRIOR APPLICATION NUMBER: 60/088028  
;; PRIOR FILING DATE: 1998-06-04  
;; PRIOR APPLICATION NUMBER: 60/088029  
;; PRIOR FILING DATE: 1998-06-04  
;; PRIOR APPLICATION NUMBER: 60/088033  
;; PRIOR FILING DATE: 1998-06-04  
;; PRIOR APPLICATION NUMBER: 60/088167  
;; PRIOR FILING DATE: 1998-06-05  
;; PRIOR APPLICATION NUMBER: 60/088202  
;; PRIOR FILING DATE: 1998-06-05  
;; PRIOR APPLICATION NUMBER: 60/088212  
;; PRIOR FILING DATE: 1998-06-05  
;; PRIOR APPLICATION NUMBER: 60/088217  
;; PRIOR FILING DATE: 1998-06-05  
;; PRIOR APPLICATION NUMBER: 60/088326  
;; PRIOR FILING DATE: 1998-06-04  
;; PRIOR APPLICATION NUMBER: 60/088655  
;; PRIOR FILING DATE: 1998-06-09  
;; PRIOR APPLICATION NUMBER: 60/088722  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088738  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088740  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088811  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088824  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088825  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088826  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088861  
;; PRIOR FILING DATE: 1998-06-11  
;; PRIOR APPLICATION NUMBER: 60/088863  
;; PRIOR FILING DATE: 1998-06-11  
;; PRIOR APPLICATION NUMBER: 60/088876  
;; PRIOR FILING DATE: 1998-06-11  
;; PRIOR APPLICATION NUMBER: 60/089090  
;; PRIOR FILING DATE: 1998-06-12  
;; PRIOR APPLICATION NUMBER: 60/089105  
;; PRIOR FILING DATE: 1998-06-12  
;; PRIOR APPLICATION NUMBER: 60/089512  
;; PRIOR FILING DATE: 1998-06-16  
;; PRIOR APPLICATION NUMBER: 60/089514  
;; PRIOR FILING DATE: 1998-06-16  
;; PRIOR APPLICATION NUMBER: 60/089538  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089598  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089653  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089908

Query Match 24.4%; Score 863.5; DB 12; Length 1894;  
Best Local Similarity 30.9%; Pred. No. 1e-44;  
Matches 325; Conservative 30; Mismatches 303; Indels 393; Gaps 35;

Qy 5 TGGQMGSGMASGRGGLGGGAGAAAAAAGGAGGGGG--LGSQGTSG----- 56  
Db 580 TGGAAAG--GTTACAGGGGATACAGTATATACATGTTCCGACTGTCAGAAAGTAGACATTG 638  
Qy 57 -RGGLGGGGAG-----AAAAAAG--GAGQGG----- 83  
Db 639 ATGATGACGAAGACCACATATCCACACAAAAATAAAAAGCTGTGAGGTTTGTATTATC 698

Qy 84 -----YGGLSQGTSGRGLGGGAGAAAAAAGGAGGGGG 124  
Db 699 CTTTGTGATGACATCTTCAAGGGAATTTAAAGGCTGAAAAAAGAGAAACAGAGGAGG 758  
Qy 125 -----LGSQGTSGRGLGGGAGAAAAAAGGAGGGGG 159  
Db 759 AAGTAAGAAATTTAAACCCAAAGGCACAAAAATTTTAGTTTACTTTCTATTGTGAGAGG 818  
Qy 160 YGGL-----GSGTSGPGG-----YGPQQTSGRGLGGGAG-----AAAAAA 199  
Db 819 AAGCTGACGAAGAGAGGAGGAGTAATTCAGATTAGTCAGAGCATGAAGGCCAAAGCA 878  
Qy 200 AAAGGAGGGYG-----GLSQGTSGRGLGGGAGAA 232  
Db 879 AAGTAGTCTACACTTGTCTTAAGGATGATCCACATCTCAGTTCTGTCTCCAGTTGTAGAA 938  
Qy 233 AAAAAAAGGAG-----OGGYGGLGSQGTSG--RGLGGGAGAAAA 272  
Db 939 GTGAAAAAGGTGATGCACCATTTAGTTGATGATGGAAGATGAAAGTGCAGAGCATG 998  
Qy 273 AAAAAAGGAGGGYGLGSQGTSGRG-----GLGGGAGAAAA--AAA 312  
Db 999 ATGAATATATTTGATGCTGATGAAAAGAACCTGATGAGAGAAAGATTTCCCAAAAAATTA 1058  
Qy 313 AAAAGGA-----GGGYGGLGSQGTSGPGYGP----- 340  
Db 1059 AAAGGACACAAGTCGAATGTTAAATCAGCTGGGAGAGGAGAGCTGAGAGAAATCAG 1118  
Qy 341 -----GQQTSGRGLGGGAGAAAAAAGGAG-----OGGYGGLGSQGTSGRGG 389  
Db 1119 TCAGCCGAGTGAAGAGCTCAGAAAGAGCAAGCAATTAANAAGCG--GNACTCTTAGCA 1177  
Qy 390 GOGAGAAAAAAGAGAGGGYGLG-----SQGTSGRGLGGGAGAAAAAAGGAG 442  
Db 1178 GCAAAACAAAAAAGTAGAAAAATCCACAAAACAGCAGAAAAAGAGTGAAGAGGAA 1237  
Qy 443 GGAG-----OGGYGGLGSQGTSGRG-----LGGGAGAAAAAAGGAG----- 483  
Db 1238 GAAGCCCTCCAGATGTTGCTGTTCCGAATACAGAGAGAAAGCAAAAGTATCAAGCT 1297  
Qy 484 -GGGYGG-----LGSQGTSGPGYGPQQTSGRGLGGGAG----- 519  
Db 1298 TTGAGGAAGCAACAGTCAAGAGAGGGAACCTTCCCGG--GAAGATCAGACCCCTTGCACTGCT 1356  
Qy 520 -----GAAAA----- 524  
Db 1357 GAACCATTTAAATCTAACTCACTCAAGCAATTTCTGAAACACCTGAAATGACATTC 1416  
Qy 525 -AAAAAGGAG--OGGYGGLGSQGTSGRGLGG-----OGAGAAAA 562  
Db 1417 TCAACAGAAAGTGAAGATGATGAAG--GATGGATGTCACATGCTACTTTCAGTTTGAGGATA 1475  
Qy 563 AAAAAAGGAGGGYGLGSQ-----TSRGG----- 589  
Db 1476 AAGCAGAAAGTGAAGATGCAAGCATGCAAGACTCAGATACATTTGAAATCTATGATC 1535  
Qy 590 --LGG-----OGAGAAAAAAGGAGGGYGLGSQGTSGRGLGGGAGAAAAAAG 642  
Db 1536 CTCGAATCCAGTGAATTAAGAGAGGAGGAGAAAGCAAAAGCTGATGAGAGAGAA 1595  
Qy 643 AAGGAGGGYGLGSQGTSGPGYGPQQT 673  
Db 1596 AAGAAA-----GAAGATAAAATGAGAATA 1619

RESULT 8  
US-09-804-898-2  
; Sequence 2, Application US/09804898  
; Patent No. US20020045264A1  
; GENERAL INFORMATION:  
; APPLICANT: DURING, MATTHEW  
; APPLICANT: XIAO, WEIDONG  
; TITLE OF INVENTION: PRODUCTION OF CHIMERIC CAPSID VECTORS

```

; FILE REFERENCE: 102182-14
; CURRENT APPLICATION NUMBER: US/09/804, 898
; CURRENT FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 60/189,110
; PRIOR FILING DATE: 2000-03-14
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 4679
; TYPE: PRT
; ORGANISM: adeno-associated virus 2
US-09-804-898-2

Query Match      20.9%; Score 742; DB 10; Length 4679;
Best Local Similarity 30.0%; Pred. No. 3.le-37;
Matches 291; Conservative 21; Mismatches 346; Indels 312; Gaps 36;

QY  2 ASMTGGQQMGRSMASGRGLGSGQAGAGAAAAA-----GGAGGGGY 46
Db  94 AGCGACGAGCGCGCAGAGAGGGAGTGGCCACTCCTACTAGGGTTCTCTGGAGGGT 153
QY  47 GGLGSGTSGRGLGCGQAGAGAAAAAAGG---AGGGYGGGSGGTSGRGLG----99
Db  154 GGAGTGGT-----GACGTGAATTACGTACATAGGCTTAG-GGAGGCTCTGATTAGAGGTCA 208
QY  100 -GGAG-----AAAAAAGAGAGGGYGGGSGTSGRGLGGG 140
Db  209 CGTAGTGTGTTTCGACATTTTCGACACCATGTGTCACGCTGGGTATTAAAGCCCGAG 268
QY  141 AGAAAA-----AAAAAAGGAG---QGGYGLGSGTSG-----PGY 174
Db  269 TGAGCAGCAGGGTCTCCATTTTGAAGCGGAGGTTTGAACGCGCAGCCGATGCCGG 328
QY  175 GPGQQTSGRGLGCGQAGAGAAAAAAGGAG-----206
Db  329 GTTTACGAGATGTGATTAAAGTCCCGACGACCTTGACGACATCTGCCGGCATTC 388
QY  207 QGGYGLGSGTSGR---GGLGCGAGAGAAAAAAGAGAGGGYGGGSGTSGRGLG 263
Db  389 TGACAGCTTTGTAAGTGGGTGG-CCGAGAAGGAATGGGAGTTGCCGCCAGATTTCTGACA 447
QY  264 GGA---GAAAAAAGAGAGG---CYGGLGSGTSGRGLGCGQAGAGAA 309
Db  448 TGGATCTGATCTGATTGAGCAGGACCCCTGACCTGCGCGAGAGCTGCACGGCGACT 507
QY  310 AAAAAAGAGGGYGGG---QGTSGPG-----GYGPGQQTSG 346
Db  508 TTCTGACGAATGGCGCGCTGTGAGTAAAGGCCCGCGAGGCCCTTTCTTTGTGCAATTTG 567
QY  347 RGGL-GGQAGN-----AAAAAAGAGGGYGG---G 376
Db  568 AGAAGGAGAGAGAGCTACTTCCACATGACGTGCTGCGTGAACCCACCGGGTGAATCCA 627
QY  377 LGSQTSRGLG-----GCGAGAAAAAAGAGAGG-----GYG 413
Db  628 TGGTTTGGGACGTTTCTCTGAGTCAGATTGCGGAAAACTGATTGAGAAATTTACCGG 687
QY  414 GLGSQ-----TSRGLGCGQAGAGAAAAAAGAGAG-GGGY 451
Db  688 GGATCGAGCGGACTTTGCCAAACTGTTCCGCGTCAACAAGACCAAGATGGCGCGGAG 747
QY  452 GLG-----SQGTSGRGLGCGQAGAGAA-----AAAA-----479
Db  748 GCGGGAACAAGTGGTGGATGAGTGTACATCCCCAATTACTTGTCTCCCAAAACCCAGC 807
QY  480 -----AGGAGGGYGGGSGTSGPGYGGQGTSGRGLG---GAGAGAA-----522
Db  808 CTGAGCTCAGTGGCGTGGACTAATATGACAGATTTTAAGCGCTCTTTGAATCTCA 867
QY  523 -----AAAAAAGAGCGGCGGGLG-----SQGTSGRGLGCGQAGAGAAAA 564
Db  868 CGGAGCGTAAACGGTTGGTGGCGGAGCATCTGACGACCATGTGCGCAGACGAGGACGAGA 927

```

RESULT 9

```

US-10-052-586-269
; Sequence 269, Application US/10052586
; Patent No. US20020127584A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C1
; CURRENT APPLICATION NUMBER: US/10/052, 586
; CURRENT FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063564
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063734
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063870
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/064103
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066120
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/066466
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066772
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/069335
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/069425
; PRIOR FILING DATE: 1997-12-12
; PRIOR APPLICATION NUMBER: 60/069870

```

; PRIOR FILING DATE: 1997-12-17  
; PRIOR APPLICATION NUMBER: 60/068017  
; PRIOR FILING DATE: 1997-12-18  
; PRIOR APPLICATION NUMBER: 60/077450  
; PRIOR FILING DATE: 1998-03-10  
; PRIOR APPLICATION NUMBER: 60/077632  
; PRIOR FILING DATE: 1998-03-11  
; PRIOR APPLICATION NUMBER: 60/077649  
; PRIOR FILING DATE: 1998-03-11  
; PRIOR APPLICATION NUMBER: 60/078886  
; PRIOR FILING DATE: 1998-03-20  
; PRIOR APPLICATION NUMBER: 60/078939  
; PRIOR FILING DATE: 1998-03-20  
; PRIOR APPLICATION NUMBER: 60/079664  
; PRIOR FILING DATE: 1998-03-27  
; PRIOR APPLICATION NUMBER: 60/079786  
; PRIOR FILING DATE: 1998-03-27  
; PRIOR APPLICATION NUMBER: 60/080107  
; PRIOR FILING DATE: 1998-03-31  
; PRIOR APPLICATION NUMBER: 60/080194  
; PRIOR FILING DATE: 1998-03-31  
; PRIOR APPLICATION NUMBER: 60/080327  
; PRIOR FILING DATE: 1998-04-01  
; PRIOR APPLICATION NUMBER: 60/080333  
; PRIOR FILING DATE: 1998-04-01  
; PRIOR APPLICATION NUMBER: 60/081049  
; PRIOR FILING DATE: 1998-04-08  
; PRIOR APPLICATION NUMBER: 60/081070  
; PRIOR FILING DATE: 1998-04-08  
; PRIOR APPLICATION NUMBER: 60/081195  
; PRIOR FILING DATE: 1998-04-09  
; PRIOR APPLICATION NUMBER: 60/081838  
; PRIOR FILING DATE: 1998-04-15  
; PRIOR APPLICATION NUMBER: 60/082568  
; PRIOR FILING DATE: 1998-04-21  
; PRIOR APPLICATION NUMBER: 60/082569  
; PRIOR FILING DATE: 1998-04-21  
; PRIOR APPLICATION NUMBER: 60/082704  
; PRIOR FILING DATE: 1998-04-22  
; PRIOR APPLICATION NUMBER: 60/082797  
; PRIOR FILING DATE: 1998-04-22  
; PRIOR APPLICATION NUMBER: 60/083322  
; PRIOR FILING DATE: 1998-04-28  
; PRIOR APPLICATION NUMBER: 60/083495  
; PRIOR FILING DATE: 1998-04-29  
; PRIOR APPLICATION NUMBER: 60/083496  
; PRIOR FILING DATE: 1998-04-29  
; PRIOR APPLICATION NUMBER: 60/083499  
; PRIOR FILING DATE: 1998-04-29  
; PRIOR APPLICATION NUMBER: 60/083559  
; PRIOR FILING DATE: 1998-04-29  
; PRIOR APPLICATION NUMBER: 60/084366  
; PRIOR FILING DATE: 1998-05-05  
; PRIOR APPLICATION NUMBER: 60/084414  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: 60/084639  
; PRIOR FILING DATE: 1998-05-07  
; PRIOR APPLICATION NUMBER: 60/084640  
; PRIOR FILING DATE: 1998-05-07  
; PRIOR APPLICATION NUMBER: 60/084643  
; PRIOR FILING DATE: 1998-05-07  
; PRIOR APPLICATION NUMBER: 60/085573  
; PRIOR FILING DATE: 1998-05-15  
; PRIOR APPLICATION NUMBER: 60/085579  
; PRIOR FILING DATE: 1998-05-15  
; PRIOR APPLICATION NUMBER: 60/085580  
; PRIOR FILING DATE: 1998-05-15  
; PRIOR APPLICATION NUMBER: 60/085582  
; PRIOR FILING DATE: 1998-05-15  
; PRIOR APPLICATION NUMBER: 60/085700  
; PRIOR FILING DATE: 1998-05-15  
; PRIOR APPLICATION NUMBER: 60/086023  
; PRIOR FILING DATE: 1998-05-18

; PRIOR APPLICATION NUMBER: 60/086392  
; PRIOR FILING DATE: 1998-05-22  
; PRIOR APPLICATION NUMBER: 60/086486  
; PRIOR FILING DATE: 1998-05-22  
; PRIOR APPLICATION NUMBER: 60/087098  
; PRIOR FILING DATE: 1998-05-26  
; PRIOR APPLICATION NUMBER: 60/087208  
; PRIOR FILING DATE: 1998-05-28  
; PRIOR APPLICATION NUMBER: 60/087609  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/087759  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/087827  
; PRIOR FILING DATE: 1998-06-03  
; PRIOR APPLICATION NUMBER: 60/088025  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088028  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088029  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088033  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088167  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088202  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088212  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088217  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088326  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088655  
; PRIOR FILING DATE: 1998-06-09  
; PRIOR APPLICATION NUMBER: 60/088722  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088738  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088740  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088811  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088824  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088825  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088826  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088861  
; PRIOR FILING DATE: 1998-06-11  
; PRIOR APPLICATION NUMBER: 60/088863  
; PRIOR FILING DATE: 1998-06-11  
; PRIOR APPLICATION NUMBER: 60/088876  
; PRIOR FILING DATE: 1998-06-11  
; PRIOR APPLICATION NUMBER: 60/089090  
; PRIOR FILING DATE: 1998-06-12  
; PRIOR APPLICATION NUMBER: 60/089105  
; PRIOR FILING DATE: 1998-06-12  
; PRIOR APPLICATION NUMBER: 60/089512  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089514  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089538  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089598  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089653  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089908

Query Match 19.48; Score 687.5; DB 12; Length 1300;  
Best Local Similarity 28.3%; Pred. No. 2e-34;  
Matches 296; Conservative 17; Mismatches 341; Indels 391; Gaps 45;



```
QY 2 ASMTGQOMGRGSMASGRGLGG--QGAGAAAAAAGAGAGQ--GGYGLGSQGT--- 54
Db 123 AGATGCTCCCTGCAGAGCGAGACTGGGCCCTCAACATGGAGATCGGCACATCAAA 182
QY 55 SGRGGLGGGAGAGAAAAAGAGGAG--OGYVGLGSQGTSGR--GLGGGAG 104
Db 183 CGAGACGGAGGAGGTCCCAAGATGCTCCGAGCAGTAAAGAAGAGATCGTGGGGA 242
QY 105 AAAAAAAGAGAGOGYVGLGSQ-----TSRGG--LGGGAGAGAAAAA 150
Db 243 TAAGAACTCCAGAGGTGATGCTGCTCTCAGTCTTAGAAACCTGTGTCAAGAACTG 302
QY 151 AGGAG-----QGYVGL-----GSQGTSGPGYGPQGTSGRGLGQ 189
Db 303 CGGGCACCCCTCCACGTGCTGTGTGCCAGCCAGGACTCGGTGGAGAGTGTGCTGTGAG 362
QY 190 G-----AGAAAA-----AAAA----- 203
Db 363 GACCATCTGCCCAAGAACACCCACCACCATCGTGCATGACAAAGTGTCAACCTCAT 422
QY 204 -----GAGGG-----YGLGSQGT----- 218
Db 423 CCAGTCTGGGTGACGCGTTCGCGAGCTCGCCGATCTGACAGGTGTGCTACCATCTA 482
QY 219 SGRGLGGGAGAA--AAAAAAGAG-----OGYVGLGSQGT--- 256
Db 483 TGAGGACCTCGGAGAGAAAGGCTGGAGTTCGCCATGACTGACCTGGACATGCTGCACC 542
QY 257 -----SRGG--LGGG-----AGA--AAAAAAGAGGGY----- 286
Db 543 CATCCACACCCAGAGGACCGTGTCAACTCAGACACATCAGACAGGATTCGTG 602
QY 287 GGLGSQ-----TSRGG----- 299
Db 603 GGCACCTGACTCCAGCAGCAAGAGGACTCTGGCCAGCAGTCTGCCCTCTGCCGCCCG 662
QY 300 -----LGGG-----AGAAAAAAGAG--OGYVGLGSQGTSGPGG 337
Db 663 CCATACTCTCCGGTGACACGCCCATAGCACCACCCGGAACAGATGGGAAGCTGCC 722
QY 338 YPGPQO--TSRGLGGGAGAAAAAAGAGOGYVGLGSQGTSG--RGLGGG--A 393
Db 723 AGTGACCTGGAGTGTGAGTGGACGTGAGGTGATGTCGGAGATGCTGACGGAGCTG 782
QY 394 GAAAAAAGAG-----AGOGYVGLGSQGTSGRGLGGGAGAAAAAAG 443
Db 783 GTGCCACCCAGCCGAGCCGACAGCTGGAGCTGCTGCAG-----GAGCTCAACGCCAC 838
QY 444 GAGOGYG--GLGSQTSRGG--LGGGAG-----AAAAAAGAG 481
Db 839 GTGCCAGCCATGCACGCGGGTCTGTAGTGATACCTGTCCGGGCCCATGCCCAAG 898
QY 482 GAG--OGYVGLG-----SQTSRGGYGPQGTSGRGLGG-----OGAGAAA 524
Db 899 GAGCCCTTCAGAGCCACACATGCCATCGAGGCTTGCTGGAGGCTGCCACAGTGGNA 958
QY 525 AAAAAAGAG-----OGYV-----GLGSQTSRGL 552
Db 959 TTCTGCCAGCCTATGTCTCCCTACCTGTCTGCTGCATGGGGCCCATGGCTTG--GCT 1017
QY 553 GG-----QGAGAAAAAAGAGOGYVGLGS-----QTSRGLGGGAGAAAAA 603
Db 1018 GGCCTAGGGTAGGGTGTGGAGGTGTGGAGGCCCTTGAGAGGCTGGCGGCCCGCAGG 1077
QY 604 AAGAGAGOGYG-----GLGSQTSRGLGGGAGAAAAA-----A 644
Db 1078 TAGRAAGTGCACACTCGCGCAGTG-----GGCAGATCTCATCAGCCCGGCTGCA 1132
QY 645 GAGOGG-----YGLGSQTSRGG 665
Db 1133 GGTGAGGCTTCAGGGGATGCTGGG 1157
```

```
RESULT 10
US-09-060-854B-2
; Sequence 2, Application US/09060854B
; Patent No. US20020081703A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David Aaron
; TITLE OF INVENTION: Human Protease and Use of Such Protease for Pharmaceutical
; TITLE OF INVENTION: Applications and for Reducing the Allergenicity of No. US2002
; FILE REFERENCE: GC532
; CURRENT APPLICATION NUMBER: US/09/060,854B
; CURRENT FILING DATE: 1998-04-15
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1497
; TYPE: PRT
; ORGANISM: B. amyloliquefaciens
US-09-060-854B-2
```

```
Query Match 19.1%; Score 675.5; DB 10; Length 1497;
Best Local Similarity 30.7%; Pred. No. 1.1e-33;
Matches 265; Conservative 22; Mismatches 330; Indels 245; Gaps 28;

QY 27 AGAAAAAAGAGAGOGYVGLGSQTSRGLGGGAGAAAAAAGAG----- 79
Db 70 AATGAAAAAAGAGAG-----GATAAAGAGTCAGAGGCCAAAAAGATATGGATCAGTTT 124
QY 80 -----GGGYVGLGSQ--GTSG-----RGL 98
Db 125 GCTGTTTGTATTAGCGTTAATCTTTACGATGGCGTTCGGCAGCACATCTCTGCCCAGGC 184
QY 99 GGGAGAAAAAAGAGAGOGYVGLGSQTSRGLGGGAGAAAAAAGAG-- 156
Db 185 GCGAGGAAATCAACCGGGAAA--AGAAATATATTGTCGGGTTTAAACACACATGAGCA 243
QY 157 OGGYG--GLGSQTSRGGYGPQGTSG-----RGLGGGAGAAAAAAGGA 205
Db 244 CGATGAGCCCGGTAAGAAGAGATGTCTTTGAAAAAGCGGGAAGTGCAAAAGC 303
QY 206 GGGYVGLGSQTSRGGG--OGAGAAAAAAGAGAGOGYVGLGSQTSRGG 263
Db 304 AATTCAAATATGTAGACGCGCTTCACTACATTAACGAA-----AAAGCTGTA--AAAG 357
QY 264 GGCAGAAAAAAGAGAGOGYVGLGSQTSRGLGGGAGAAAAAAGAGGAG 323
Db 358 AATTGAAAAAGACCGCGCTGCTTACCTTGAAGAG--ATCAGTAGCACATGCGT 414
QY 324 YGLGSQTSRGGYGPQGTSGRGLGGGAGAAAAA----- 364
Db 415 AGCGCAGTC---CGTGCTTACGCGGTATCAAAATTAAGCCCTGCTCTGCACCTCT 470
QY 365 --AAGG---AGOGYVGLGSQTSRGLGGGAGAAAAAAGAG----- 408
Db 471 CRAAGGTACACTGGATCAATGTTAAGTAGCGGTTATCGACAGCGGTATCTCT 530
QY 409 -----OGGYVGLGSQTSR--RGLGGG----- 430
Db 531 CATCTCTGATTAAAGGTAGCAGCGGAGCGAGCATGTTCTTCTTGAACAAATCTCTTC 590
QY 431 --AGAAAAA---AAAGGAG-----OGGYVGLGSQTSRGLGGGAGAAAAA 477
Db 591 CAAGAGCAACAACCTCTCACGGAACCTACGTTGCCGCGACAGTTTCGCGCTCTTAATAACTCA 650
QY 478 AAGAGAG-----OGYV-----GLGSQTSRGG-----YGPQGT 508
Db 651 ATCGGTGTATTAGCGCTTGGCCAGCGCATCACTTTACGCTGTAAAGTCTTCGGTGT 710
QY 509 SGRGLGGGAGAAAAAAGAGAGOGYVGLGSQTSRGLGGGAG--GAAAAA 566
Db 711 GACGGTTCGGGCAATACAGCTGATCATTAACGGAATCGAGTGGGCGGATCGCAACAT 770
```



Db 730 CAGACCTATGAGCTAACGAGACAGCAAGCAGCATGAGTGTATGACACTGGATGGCTTC 789  
QY 247 GYGLGS---QGTSGR---GGLGGOGA-----GAAAAAAAAGGAGGOGYGLGS 291  
Db 790 ATGATGTAAGTGTGTTCGGCGGAGGGGACTGCTTGGACAACACCCACACGT-----GTGT 845  
QY 292 QGTSRGLGGOGAGAAA-----AAAAA----- 315  
Db 846 GTTCCAGGACATGAACACAGCCCTTCCCACTACTTCACTCTCTCTCCCAACACCTA 905  
QY 316 -----AGAGOGYGLGSOGTSPGPGYGPQOQTSRGRGLGGOGAGAAAA 361  
Db 906 TCTGACTGACTCCAGATGGGGGGCCAGCAGCAGCCAGGCGCTATGTTAGGCGCTTTC 965  
QY 362 AAAAAAGAGOGYGLGSOGTSGRGLGGOGAGAAAAAAGAGGAGGOGY----- 413  
Db 966 CAGGGATCGCGTGGAGTGGAGTGTCTGGAGGGGGCCAGAGGGGCGCCCTCAT 1025  
QY 414 -----GLGSQGTSG-----RGGLGGGAG 432  
Db 1026 CTATCATGCCATPACCTCACCTCCAAGATTCTTCCGGGAGCTGCCCAAGCCGTGCG 1085  
QY 433 -----AAAAAAGAGGAGQ 447  
Db 1086 CGACCATGCTTCACGCTGTCCCTTACCTGTCTATCTCTATCCCTGGAGAACCACTGCGG 1145  
QY 448 GYGGLGSOGTSGRGLGGOGA-----GAAAAAAGAGG-----QGGYG 489  
Db 1146 GCTGGAGCAGCA--GGCTGCCATGGCCGCCACCTCTGCACCATCTCTGGGGGACATGCTG 1203  
QY 490 GLGSOGTSPGPGYGPQOQ-----SGRGLGGOG-----AGAAAAAAGAG 531  
Db 1204 GTGACACAGCGCTGACTCCCAATCCCGAGGAGCTGCCATCCCAAGAGCTGAAG 1263  
QY 532 GAGOGYGLGSOGTSGRGLGGOGAGAAAA-----AAAAAGAGOGYGLGS 581  
Db 1264 GGCCTGTGTAAGGGAAGAGCTGCCCGCTGCTCGGAGCAGGATGG 1316  
QY 582 -----QGTSG-----RGLGGOGAGAAAAAAGAGGOGYGG-LGSQTSRGLG 628  
Db 1317 CCGGGCTGTGCGATCGGAGGAGGGGAGGAGGATGACGAGGAGAGAGAGAGGT 1376  
QY 629 GGAGAGAAAAAAGAGGAGG-----YGLGSQGTSG-----PG 664  
Db 1377 GGAG-CCTCAGCGCAGAGCGGCTGCCCAAGCAGATCTCCCGGAGCTGTCCGCCCTG 1435  
QY 665 GPGPGQQTSG 674  
Db 1436 CTGTGTACTG 1445

RESULT 13  
US-09-879-957-37  
; Sequence 37, Application US/09879957  
; Patent No. US20020034755A1  
; GENERAL INFORMATION:  
; APPLICANT: SPARKS, Andrew B.  
; HOFFMAN, No. US20020034755A1h  
; KAY, Brian K.  
; FOWLES, Dana M.  
; McCONNELL, Stephen J.  
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
; DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
; USING SAME  
; NUMBER OF SEQUENCES: 227  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/879,957  
FILING DATE: 13-Jun-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/630,915  
FILING DATE: 03-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Misrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-174  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1400 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 37:  
US-09-879-957-37  
  
Query Match 18.4%; Score 652; DB 10; Length 1400;  
Best Local Similarity 28.3%; Pred. No. 2.7e-32;  
Matches 274; Conservative 24; Mismatches 325; Indels 344; Gaps 32;  
  
QY 16 ASRGGGLGGOGAGAAAAAAGAGGAGGOGYGLGSOGTSGRGLGGOGAGAAAAA 75  
Db 2 AAAGGAGGAGAGTGTCAAAA--GAAGGATGGCA-----GGAAAAAGCAACA 50  
QY 76 AGAGOGYGLGSOGTSGRGLGG-----QGAGAAAAAAGAGGAG 120  
Db 51 GGAAGCAGAGCAGAGCTGGTGGCTTTTCCATCAACACCAAGAACCACTAGGCA-- 108  
QY 121 GYGLGSOGTSGRGG-----LGGO-----GAGAAAA----- 147  
Db 109 -----GCTGTCAGGACCCCTGCTCCACTCCAGAAAAAGGTCCACTTACCTTTCTGCA 163  
QY 148 -----AAAAAGAGGOGYGLGSOGTSGPG-----YGPQOQTSRGG---L 186  
Db 164 CAGGAAATGTAAGTGTGTTATACCGGCGACTGTACCCCTTTGAAATCCAGAGCCAT 223  
QY 187 GGOGAGAAAAAAGAGGAGGOGYGLGSOGTSGRGLGGOGAGAAAAAAGAGGAG 246  
Db 224 GATGAATCACTATCAGCCAGGAGACATAGTCATGTTGGATGAAAGCAAACTGGAGAA 283  
QY 247 GYGLGSOGTSGRGLGGOGAGAAAAAAGAGGOGYGLGSOGT-----SGRG 298  
Db 284 CCGG-CCTGGCTTGGAGGAGAAATTAAGGAAGACAGG-----GTGGTTCCCTGCAACT 338  
QY 299 GLGGOGAGAAAAAAGAGGOGYGLGSOGTSGPGYGPQOQTSRGLGGOGAGAA 358  
Db 339 ATCAGAGAAAAATCCCAAGAAATGAGTTCCCGCTCCAGTGAACCAAGTCACTGATTC 398  
QY 359 AA-----AAAAAGGAGGOGYGLGS-----OGT----- 382  
Db 399 CATCTGCCCTTGCCTCCCAAACTGGCTTCCGTGAGACCCCGCCCTTTTGGCAGTAACT 458  
QY 383 -----SGRG-----GLGG----- 390  
Db 459 CTCAGAGCCCTCCAGACCCCTTAATACTGGCCCACTTCACTCCAGCTGCCCAACCA 518  
QY 391 -----QGAGAAAAAAGAGGOGYGLGSOGTSGRGLGG---QGAGAAAAA 440  
Db 519 GCACGAATGAGAAACCAAGAACGATAACTGGGATGATGGGCGAGCCCGCTCTCTCA 578

Qy 441 AAG-----GAGGGYGGGSGQTSRGGGCGG-----AGAAAAA--- 479  
Db 579 CCGTTCCAAGTGCAGGC-CAGTTAAGCAGAGGTCGCGCTTTACTCCAGCCAGCGCCACT 637  
Qy 480 -----AGGAGGGYGGGSGQTSRGGGCGG--- 513  
Db 538 GGCTCTCCCGTCTCTGCTAGCCAG--GGTAAAGGTCGAGGGGCTACAAGCTC 695  
Qy 514 -----LGGGAGAAAAA----- 530  
Db 696 AAGCCCTATATCTTGGAGACCAAAAGACACACCTTAATTTTAAACAAATGATG 755  
Qy 531 -----GGAGGGYGGGSGQTSRGGGCGGAGAA---AAAAAAGAGG 573  
Db 756 TCATCACCGTCTCGAACAGACATGTGTGTTGGAGAAAGTTCAAGGTGAGAGG 815  
Qy 574 GGYGGLG---SOGT-----SGRGLGGGAGAAAAAAGGAG 610  
Db 816 GTTGGTCCCAAGTCTTACGTGAACATTCAGGCCCCCATAGGAAAGTCTACAAGCA 875  
Qy 611 QGG----YGGGSGQTSRGG---LGGGAGAAAAAAGGAG-----QGGYGGG 657  
Db 876 TGGATTCTGTTCTCAGAGAGTCTGCTAGCTTAACCGAGTAGCCTCTCCAGCAGCA 935  
Qy 658 QGTSGP 664  
Db 936 AGCCGAG 942

RESULT 14

US-09-756-071B-20  
; Sequence 20, Application US/09756071B  
; Patent No. US20020052307A1  
; GENERAL INFORMATION:  
; APPLICANT: Trygsvason, Karl  
; Kallunki, Pekka  
; Pyke, Charles  
; TITLE OF INVENTION: Laminin Chains: Diagnostic Uses  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fay Sharpe Fagan Minnich & McKee  
; STREET: 1100 Superior Ave, Suite 700  
; CITY: Cleveland  
; STATE: Ohio  
; COUNTRY: USA  
; ZIP: 44114  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/756,071B  
; FILING DATE: 08-Jan-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 09/663,147  
; FILING DATE: 150-September 2000  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Minnich, Richard, J.  
; REGISTRATION NUMBER: 24,175  
; REFERENCE/DOCKET NUMBER: TRV 20014  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 216-861-5582  
; TELEFAX: 216-241-1666  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 720 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:  
US-09-756-071B-20

Query Match 18.4%; Score 651.5; DB 10; Length 720;  
Best Local Similarity 32.5%; Pred. No. 1.7e-32;  
Matches 249; Conservative 17; Mismatches 314; Indels 187; Gaps 35;  
Qy 16 ASRGRGLGGGAGAAAAAAGAGGGYGGGSGQTSRGGGCGGAGAAAAA--- 75  
Db 8 ATAGGGAGGTGGCCAGTCAATAGT-----TACTTTATGATTG-----CTAACCC 54  
Qy 76 AGGAGGGYGGGSGQTSRGG---LGGGAGAAA-----AAAAAAGGAGGGYGG 124  
Db 55 TGGTGAGCAGGAAGTATGTGGACAGAGAGAAACCCCTTGGTTCAGCCTGGAGAAA-GG 113  
Qy 125 LGSQTSRGGGAGAAAAAAGAGGGYGGGSGQTSRGGGCGGAGCCCTGTTGTGACTCTC 172  
Db 114 AGAGGTT-----GACCCCTAACTGGAG-GGTGGAGAGGCCCTGTTGTGACTCTC 162  
Qy 173 -GYPGQQTSG-----RGLGGGAGAAAAAAGAGGGYGGGSGQTSRGG-L 224  
Db 163 CGACTGACTTCTTCTTCTTATG-----TCCTTTAAGCCGGA-----GCTGATTC 207  
Qy 225 GSGAGAAAAAAGAGGGYGGGSGQTSRGGGCGGAGAAAAA-----AAAAAG 279  
Db 208 GGGTGTGCTTCTTCTTCTTATGTTAGCTTGGGCTTCCAGTTTGGAGGAG 267  
Qy 280 CAGGGYGGGSGQTS-----SGRGLGGGQ-----AGAAAAAAGAGGGYGGG 329  
Db 268 GGGCGG-GCTGCTGTCTACCTCTGTGATCTGCCCTGGACCCCGGAGAGGAGG 326  
Qy 330 QTSRGGYGGGSGQTSRGGGCGGAGAAAAAAGAGGGYGGGSGQTSRGG-L 388  
Db 327 -----GCTCCGGGGAATCTCGCACATTCAGGCAAGGCTCCGGGCGGCGAGCC 375  
Qy 389 GGGAGAAAAAAGAGGGYGGGSGQTSRGGGCGGAGAAAAAAGAGGGYGGGAGG 448  
Db 376 TCTGTGCCACACCTTGGCCCG--GGCCAGGT-----GTGCGGCTCTCTCGCTGCCAG 426  
Qy 449 YGGG-L-SQGTSGRGLGGGAGAAAAAAG---GAGGGYGG---GLGSGQTSG-- 498  
Db 427 GGGGAGCGGGCGGCTGCGGGGAGCGATTTCCAGCCCGGTTTGTGCTGTGTGTTGTC 486  
Qy 499 -----PGYPGQQTSGR-----GGLGGGAGAAA-----AAAAAAG-- 532  
Db 487 TGCCTCTGGAGGGCTGGGCTCTCTCTTATTCACAGGTAGTACACCTTGAACACAGGCT 546  
Qy 533 -----AGGGYGGG--SQGTSRGGGCGGAGAAAA-----AAAAAGGAGG 573  
Db 547 CTCCTCTCTCAG-GACTGAGTCAAGTAGAGATCGATAAAACACCTGATCAAGGAAA 605  
Qy 574 -GGYGGGSGQTSRGG-GLGGGAGAAAAAAGAGGGYGGGSGQTSRGGG 630  
Db 606 AGGAAGCACAGCGGAGCGGAGGTGAGAACCAACCAACCGAGGCGCG------GGCAG 658  
Qy 631 QCA-----GAAAAAAGAGGGYGG---YGGGSGQTSGP 664  
Db 659 CGACCCCTGACGCGGAGACAGAGACTGAGCGGCGCGGCGGCGCCCATG 705

RESULT 15

US-10-052-586-525  
; Sequence 525, Application US/10052586  
; Patent No. US20020127584A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.

APPLICANT: Zhang,Zemin	PRIOR FILING DATE: 1998-04-08
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC	PRIOR APPLICATION NUMBER: 60/081199
TITLE OF INVENTION: ACIDS ENCODING THE SAME	PRIOR FILING DATE: 1998-04-09
FILE REFERENCE: P3430R1C1	PRIOR APPLICATION NUMBER: 60/081838
CURRENT APPLICATION NUMBER: US/10/052,586	PRIOR FILING DATE: 1998-04-15
CURRENT FILING DATE: 2002-01-15	PRIOR APPLICATION NUMBER: 60/082568
PRIOR APPLICATION NUMBER: 60/059263	PRIOR FILING DATE: 1998-04-21
PRIOR FILING DATE: 1997-09-18	PRIOR APPLICATION NUMBER: 60/082569
PRIOR APPLICATION NUMBER: 60/059266	PRIOR FILING DATE: 1998-04-21
PRIOR FILING DATE: 1997-09-18	PRIOR APPLICATION NUMBER: 60/082704
PRIOR APPLICATION NUMBER: 60/062250	PRIOR FILING DATE: 1998-04-22
PRIOR FILING DATE: 1997-10-17	PRIOR APPLICATION NUMBER: 60/082797
PRIOR APPLICATION NUMBER: 60/063120	PRIOR FILING DATE: 1998-04-22
PRIOR FILING DATE: 1997-10-24	PRIOR APPLICATION NUMBER: 60/083322
PRIOR APPLICATION NUMBER: 60/063121	PRIOR FILING DATE: 1998-04-28
PRIOR FILING DATE: 1997-10-24	PRIOR APPLICATION NUMBER: 60/083495
PRIOR APPLICATION NUMBER: 60/063486	PRIOR FILING DATE: 1998-04-29
PRIOR FILING DATE: 1997-10-21	PRIOR APPLICATION NUMBER: 60/083496
PRIOR APPLICATION NUMBER: 60/063540	PRIOR FILING DATE: 1998-04-29
PRIOR FILING DATE: 1997-10-28	PRIOR APPLICATION NUMBER: 60/083499
PRIOR APPLICATION NUMBER: 60/063541	PRIOR FILING DATE: 1998-04-29
PRIOR FILING DATE: 1997-10-28	PRIOR APPLICATION NUMBER: 60/083559
PRIOR APPLICATION NUMBER: 60/063544	PRIOR FILING DATE: 1998-04-29
PRIOR FILING DATE: 1997-10-28	PRIOR APPLICATION NUMBER: 60/084366
PRIOR APPLICATION NUMBER: 60/063564	PRIOR FILING DATE: 1998-05-05
PRIOR FILING DATE: 1997-10-28	PRIOR APPLICATION NUMBER: 60/084414
PRIOR APPLICATION NUMBER: 60/063734	PRIOR FILING DATE: 1998-05-06
PRIOR FILING DATE: 1997-10-29	PRIOR APPLICATION NUMBER: 60/084639
PRIOR APPLICATION NUMBER: 60/063870	PRIOR FILING DATE: 1998-05-07
PRIOR FILING DATE: 1997-10-31	PRIOR APPLICATION NUMBER: 60/084640
PRIOR APPLICATION NUMBER: 60/064103	PRIOR FILING DATE: 1998-05-07
PRIOR FILING DATE: 1997-10-31	PRIOR APPLICATION NUMBER: 60/084643
PRIOR APPLICATION NUMBER: 60/065311	PRIOR FILING DATE: 1998-05-07
PRIOR FILING DATE: 1997-11-13	PRIOR APPLICATION NUMBER: 60/085573
PRIOR APPLICATION NUMBER: 60/066120	PRIOR FILING DATE: 1998-05-15
PRIOR FILING DATE: 1997-11-21	PRIOR APPLICATION NUMBER: 60/085579
PRIOR APPLICATION NUMBER: 60/066466	PRIOR FILING DATE: 1998-05-15
PRIOR FILING DATE: 1997-11-24	PRIOR APPLICATION NUMBER: 60/085580
PRIOR APPLICATION NUMBER: 60/066772	PRIOR FILING DATE: 1998-05-15
PRIOR FILING DATE: 1997-11-24	PRIOR APPLICATION NUMBER: 60/085582
PRIOR APPLICATION NUMBER: 60/069335	PRIOR FILING DATE: 1998-05-15
PRIOR FILING DATE: 1997-12-11	PRIOR APPLICATION NUMBER: 60/085700
PRIOR APPLICATION NUMBER: 60/069425	PRIOR FILING DATE: 1998-05-15
PRIOR FILING DATE: 1997-12-12	PRIOR APPLICATION NUMBER: 60/086023
PRIOR APPLICATION NUMBER: 60/069870	PRIOR FILING DATE: 1998-05-18
PRIOR FILING DATE: 1997-12-17	PRIOR APPLICATION NUMBER: 60/086392
PRIOR APPLICATION NUMBER: 60/068017	PRIOR FILING DATE: 1998-05-22
PRIOR FILING DATE: 1997-12-18	PRIOR APPLICATION NUMBER: 60/086486
PRIOR APPLICATION NUMBER: 60/077450	PRIOR FILING DATE: 1998-05-22
PRIOR FILING DATE: 1998-03-10	PRIOR APPLICATION NUMBER: 60/087098
PRIOR APPLICATION NUMBER: 60/077632	PRIOR FILING DATE: 1998-05-28
PRIOR FILING DATE: 1998-03-11	PRIOR APPLICATION NUMBER: 60/087208
PRIOR APPLICATION NUMBER: 60/077649	PRIOR FILING DATE: 1998-05-28
PRIOR FILING DATE: 1998-03-11	PRIOR APPLICATION NUMBER: 60/087609
PRIOR APPLICATION NUMBER: 60/078886	PRIOR FILING DATE: 1998-06-02
PRIOR FILING DATE: 1998-03-20	PRIOR APPLICATION NUMBER: 60/087759
PRIOR APPLICATION NUMBER: 60/078939	PRIOR FILING DATE: 1998-06-02
PRIOR FILING DATE: 1998-03-20	PRIOR APPLICATION NUMBER: 60/087827
PRIOR APPLICATION NUMBER: 60/079664	PRIOR FILING DATE: 1998-06-03
PRIOR FILING DATE: 1998-03-27	PRIOR APPLICATION NUMBER: 60/088025
PRIOR APPLICATION NUMBER: 60/079786	PRIOR FILING DATE: 1998-06-04
PRIOR FILING DATE: 1998-03-27	PRIOR APPLICATION NUMBER: 60/088028
PRIOR APPLICATION NUMBER: 60/080107	PRIOR FILING DATE: 1998-06-04
PRIOR FILING DATE: 1998-03-31	PRIOR APPLICATION NUMBER: 60/088029
PRIOR APPLICATION NUMBER: 60/080194	PRIOR FILING DATE: 1998-06-04
PRIOR FILING DATE: 1998-03-31	PRIOR APPLICATION NUMBER: 60/088033
PRIOR APPLICATION NUMBER: 60/080327	PRIOR FILING DATE: 1998-06-04
PRIOR FILING DATE: 1998-04-01	PRIOR APPLICATION NUMBER: 60/088167
PRIOR APPLICATION NUMBER: 60/080333	PRIOR FILING DATE: 1998-06-05
PRIOR FILING DATE: 1998-04-01	PRIOR APPLICATION NUMBER: 60/088202
PRIOR APPLICATION NUMBER: 60/081049	PRIOR FILING DATE: 1998-06-05
PRIOR FILING DATE: 1998-04-08	PRIOR APPLICATION NUMBER: 60/088212
PRIOR APPLICATION NUMBER: 60/081070	PRIOR FILING DATE: 1998-06-05



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:13:49 ; Search time 20.4171 Seconds  
(without alignments) 3253.588 Million cell updates/sec

Title: US-09-490-291-6

Perfect score:

Sequence: 1 MASMTCGQQMGRGSMASGRG.....TSGIRAPSTSEHHHHHHH 691

Scoring table: BLOSUM62

scoring cadre. DEOSOM02  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 2832224

Minimum DB seq length: 0

Minimum	DB seq	length:	0
Maximum	DB seq	length:	2000000000

Post-processing: Minimum Match 08

Maximum Match 100%

## Listing first 45 summaries

Database :

```
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query %		DB	ID	Description
		Match	Length			
1	2138	60.3	718	2	A36068	major ampullate fi
2	1386	39.1	2639	2	T31328	fibroin - Chinese
3	1281.5	36.2	1901	2	F70806	hypothetical glyci
4	1260.5	35.6	1489	2	F70807	hypothetical glyci
5	1242	35.1	1079	2	B70807	hypothetical glyci
6	1198.5	33.8	1306	2	A70934	hypothetical glyci
7	1196.5	33.8	1381	2	E70806	hypothetical glyci
8	1180	33.3	1660	2	A70869	hypothetical glyci
9	1158.5	32.7	1538	2	F70846	hypothetical glyci
10	1152	32.5	1329	2	E70917	hypothetical glyci
11	1151	32.5	853	2	A70896	hypothetical glyci
12	1138.5	32.1	627	2	A44112	spidroin 2, dragli
13	1114.5	31.5	749	2	A70812	hypothetical glyci
14	1110.5	31.3	914	2	F70987	hypothetical glyci
15	1083	30.6	882	2	B70812	hypothetical glyci
16	1049.5	29.6	778	2	F70963	hypothetical glyci
17	1038.5	29.3	767	2	E70895	hypothetical glyci
18	1032.5	29.1	837	2	E70835	hypothetical glyci
19	1032	29.1	957	2	F70835	hypothetical glyci
20	1014	28.6	783	2	E70824	hypothetical glyci
21	1007	28.4	731	2	C70974	hypothetical glyci
22	992	28.0	714	2	A70807	hypothetical glyci
23	989	27.9	741	2	F70917	hypothetical glyci
24	988.5	27.9	801	2	F70824	hypothetical glyci
25	977.5	27.6	1011	2	F70620	hypothetical glyci
26	976	27.5	860	1	EAMS	elastin precursor
27	939	26.5	864	1	EART	elastin precursor
28	937	26.4	923	2	F70820	hypothetical glyci
29	923.5	26.1	567	2	A70893	hypothetical glyci

## ALIGNMENTS

## RESULT 1

A36068

major ampullate fibroin protein - orb spider (*Nephila clavipes*) (fragment)

C; Species: *Nephila clavipes*

C;Date: 08-Mar-1991 #sequence\_revision 13-Jan-1993 #text\_change 09-Sep-1997

C;Accession: A36068

R; Xu, M.; Lewis, R. V.

Proc. Natl. Acad. Sci. U.S.A. 87, 7120-7124, 1990

A;Title: Structure of a protein superfiber: spider drag

**A; Reference number:**

A;Accession: A36068

A;Status: preliminary;

A;Molecule type: DNA

A;Residues: 1-718 <XUA>

A;Cross-references: GB:M3/I37; NID:gl59711; PID:gl59712

Query Match  
60 38: score 2138. DB 2: Length 718.

Query match	60.3%;	Score 2138;	DB 2;
Best Local Similarity	66.3%;	Pred. No. 3	6e-113;

Best Local Similarity	60.3%, F10: NO: 3.0E-113;
Matches 498: Conservative	20: Mismatches 49: Indels 184: Gaps 35:

Qy	25	QGAGAAAAAAGGAGGGYGGGLGSGQT--SGRGGLGCGAG-AAAAAAGAGGAGG	81
Db	1	QGAG----AAAAAAGGAGGGYGGGLGCGAGGCGYGGGLGCGAGGAGAAAAAGAGG	56
Qy	82	GGYGGGLGSGQTSGRGGGLGGGAGAAAAAAGGAGGAGCGYGGGLGSGQTSGRGGGLGCGA	141
Db	57	GGYGGGLGSGQ-----AGRGGCGAG-----AAAAAAGGAGGCGYGGGLGSGQ-AGRGGGLGGGA	107
Qy	142	GAATAAAAAAAGGAGGCGYGGGLGSGQTSGPGYGPQQTSGRGGGLGGCGAGAAAAAAGG	201
Db	108	G-----AAAAAAGGAGGCGYGGGLGSGQ-----AGRGGCG-----AAAAA	142
Qy	202	AGGAGCGGYGGLGSGQTSGRGGGLGGGAGAGAAAAAAGG-----AGGAGCGGYGGGLG	252
Db	143	AGGAGCGGYGGLGSGQ-AGRGGGLGGCGAGAAAAAAGGAGGCGYGGGLGGCGAGGCGYGGGLG	201
Qy	253	SGQTSGRGGGLGGGAGAAAAAAGG-----AAAAAGAGCGGYGGGLG	291
Db	202	SGQ-AGRGGGLGGGAGAAAAAAGGAGGAGCGGLGCGAGCGAGAGAAAGGAGCGGYGGGLG	260
Qy	292	QGTSGRGGGLGGCGAGAAAAAAGGAGGCGYGGGLGSGQTSGPGYGP-GQQTSGRGGGL	350
Db	261	QG-----AGRGGEGAG-----AAAAAAGGAGCGGYGGGLGGGQ-AGCGGYGGGLGSGAGRGGGL	311
Qy	351	GGGAGAGAAA-----AAAAAAGGAGCGGYGGGLGSGQTSGRGGGLGGCG	392
Db	312	GGGAGAAAAAGGAGCGGLGGCGAGCGAGAAAAAAGGAGGCGGYGGGLGSGQ-AGRGGGLGGCG	370
Qy	393	AGAAAAAAGGAGGCGGYGGGLGSGQTSGRGGGLGGGAGAAAAAAGGAGGAGCGGYG	452
Db	371	AG---AVAAAAAGGAGCGGYGGGLGSGQ-----ACRGGCGAG-----AAAAAAGGAGCGGYG	419





Db	340	GDVGLGVGGTGGKGVGGVAGLGGAGGAAGCOLFSAGGAAGAAGVGGTGGCGGAGGACAA	399
Qy	91	G-----TSRGGGLGGCGACAAAAAAGAGGAGCGGCGGGLGSQGTSGR--	133
Db	400	GADAPASTGLTGTGFAGCAGGVGGGGNAIAGGINGSAGAGGTGGCGAGGMSGADN	459
Qy	134	-----GGLGGG--AGAAAAAAAAGG-----ACGGYGGGLGSQGT	168
Db	460	ASGIGADGGAGGTGGNAGAGGAGGAAGTGTGGVYGAAGKAGIGTGGCGGAGGAGSAGT	519
Qy	169	-----SPGGYGPQQTSRGGGLGGCGAGAAAAAAAAGGAGGQ-----	207
Db	520	DATATGATGCTGFSGCAGCAGCAGGNTGVGGTNGSSGGQ--GTGGAGGAGGAGVGDNP	578
Qy	208	--GGYGLGSQGTSGRGLGGGQ--AGAAAAAAAAGG--ACGGYGGGLGSQGT---SGR	259
Db	579	GIGTGTGTGKG--GAGGAGGGSSGAGGTNGSGGAGTGGCGAGGAGGAGADNPTGI	636
Qy	260	GGGLGGCGAGAAAAAAAAGGAGCGGCG-------LGSQGTSGRGLGGQ--GACAAAA	311
Db	637	GGAGGTGTTGGAAGAGGAGGAGGAGTGGTGGVAGSVGNAGIGTGTGGVGGAGGAGNA	696
Qy	312	AAAAAGGA--CGGYYGGLGQ---GTSGGPGYGPQQTSRGGGLGGCGA-----	355
Db	697	GSSNTGGAGPAGCAGGCGGAGGNSGVGGTNGSSGAGGAGGKGTGGAGGCSADNP	756
Qy	356	-----GAAAAAAAAGGA--GCGYGGGL-----GSQGTSGRGLGGQGA-----	393
Db	757	AGGAGGTGGAAGAGGAGGATGTGTTGGVVGATGSAGIGGAGRGSGDGDGASGLGLG	816
Qy	394	-----CAAAAAAAAAGGA--GCGYGGGLGSQGTSG---RGLGGQ--GAGAAA--	435
Db	817	FDGGGGGGGAGGSAGGAGGAGGAGGNGSDGSDGATGAAGLDGNGVGGVGGGAGGAG	876
Qy	436	-----AAAAAAGGAGCGYGGGLGSQG---TSRGGGLGGCGA-----GAAAA	475
Db	877	NGGNAGVGLTAKAGDCAAGNCGGAGCAGCAGGAGDNNNGCGAGCGCGCGGLGCA	936
Qy	476	AAAAAGGA-----CGGYYGGLGSQGTSGPGGYGPGQQTSGRGLGGQGAGAAAAA	531
Db	937	SINANGAGGNGCTGCKGAGGAGTLLGVGSG---GTGGDGDAGSGGGGFGGAAGKAG	993
Qy	532	GACGGYGGGLGSQGT'S-----GHRGLGGCGAGAAAAAAAAGCAGCGGGLG	580
Db	994	GGNGRGSDGSDASGLGLGLSDFDGGQGGGAGGSAGAGGTNGAGAGGNGGDDGDG	1053
Qy	581	SQGTSG---RGLGGQ--GAGAAA-----AAAAAAGGAGCGYGGGLGSQGT---	621
Db	1054	ATGAAGLCDNGVGGDGGAGGAGNCGNAGVGLTAKAGDGGNAGNCGNAGGAGGAG	1113
Qy	622	--TSRGGGLGGQGA-----GAAAAAAAAGGA-----GCGYGGGLGSQGTSGPGY	667
Db	1114	NFNGCGGAGGCGCGGLGAGATT'SINANGAGGNGGTGGKGGAGGACTLGVGSG	1169

RESULT 4

D70807

hypothetical glycine-rich protein Rv3514 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C:Accession: D70807

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: D70807

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1489 <COL>

A:Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAAL17751.1; PID:g292445

A; Experimental source: strain H37Rv  
C; Genetics:  
A; Gene: Rv3514  
C; Superfamily: collagen alpha 1(IV) chain

Query Match	35.6%	Score 1260.5;	DB 2;	Length 1489;
Best Local Similarity	38.5%;	Pred. No. 7.1e-64;		
Matches 344;	Conservative	35;	Mismatches 274;	
			Indels 241;	Gaps 37;

Qy	6	GGQQMGRGSMASGRGLGQ-----GAGAAAAA-----GAGAAAAA-----AA	35
Db	214	GGGAGGCGGAGCGCTGGAGGABELLFCAGCAGGAGTDCGPCATCCTGGHGVGGCGWLAP	273
Qy	36	AAAGCA-CQGGYGGGLSQG---TSRGGGLGQ-CAGAAAAAAGAGAGQGGYGGGLGQ-90	
Db	274	GGAGGAGGCGGAGGSDGALGCTGCTGCAGCAGGAGGALLLGGAGGQGGGLGAGGQG	333
Qy	91	-----GTSGRGLGQ-----QGAAAAAAGAGAGQGGYGGGLSQGQTS	131
Db	334	GTGAGGCGVLYGVGTGKGVGAGLGGAGCAAGQLFSASGAAGNAGVCGACGCG-391	
Qy	132	GRGLGQGA-----GAAAAAAGGA-----GQGYGGLSQGTSRGGY 175	
Db	392	GDGAGCAGADADPCATGCTGFAGCAGCAGGAGSSGAGCTNGSGAGGCGGACGCA	451
Qy	176	PCQOTSRRGLGCGAGAAAAAAGAGAGCGGYG-LGSQTSRRGLGQ-----227	
Db	452	GADNPTGIGTGDDGTGGAAGAGGAGGAAGTGGTGGMTGTTGNAGVGGAGGQGGGAG	511
Qy	228	GAGAAA-----AAAAAAGCA-----GQGYGGLSQG-----255	
Db	512	GAGADADPCATGTFAGCAGCAGGAGSSGAGCTNGSGAGGTGGGAGCAGCAGAD	571
Qy	256	-TSRRGLGCGGAGAAAAAAGAGAGCGGYG-LGSQTSRRGLGQ-----GAGAAA	309
Db	572	NPTGIGTGCGDGTGGAAGAGGAGGAAGTGGTGGMTGTTGNAGVGGAGGCGGCGAG	631
Qy	310	AAA-----AAAGGAGQ-----GGYGGLSQ-GTSGPGGYGPGQOT	344
Db	632	ADADPCATGCTGFAGCAGCAGGAGGAGSSAGCTNSSGSAGTGRSGTGGAGGAGDNPT	691
Qy	345	SGRGLGCGCAGAAAAAAGAGCAGCGYGG-LGSQTSRRGLGCGCAGAAAAA	403
Db	692	-GIGTGCGDGTGGAAGAGGAGGAAGTGGTGGMTGTTGNAGVGGAGGSSGAGTNGSGA	750
Qy	404	AGGAGCGGYGGLSQGT---SRRGLGCGCAGAAAAAAGAGAGCGGYG-LGSQTS	459
Db	751	GGTDGCGCAGGAGGADNPTGICGCTGGCGCTCGAAGCAGCAGAGTGGTGMIGTTGNA	810
Qy	460	GRGLGQ-----GAGAAAAA-----AAAAAG-----483	
Db	811	GVGGAGCGGDDGAGGAGADAPGATGTFAGGAGGAGSGGSSCAGTNGSGAGGT	870
Qy	484	-----CQGYGGLSQGTSRGG-LGCGCAGGAGAA-----523	
Db	871	CGQVAVGAGISFSGNSNGCTGCTGGV--GTGGDGNAGTCAGDPGKGTGCTG-CTGC	927
Qy	524	AAAAAAGGAG-CQGYGGLSQGTSRRGLGCGCAGAAAAAAGAGAGCGGYGGGLGQ	582
Db	928	SCGAGGSGGANFNGTGGT--CTGCKGGLNTDGLSSATGCTGCTGCTGCKCTCGACDD	985
Qy	583	TSRRGLGCGGAGAAAAAAGGA-----GQGYGGLSQGTSRRGLGQ-----631	
Db	986	SAGTGTGCGAGGAGAGGLANTGTTAGNAGTGGDGGGNGGQDSDG-SGLGGPGGPFAG	1044
Qy	632	-----CAGAAAAA-----AAAAAGAGCGGYGGL-----SQTSGPGY 667	
Db	1045	GAGCKGAGSSGAGTNGSGAGGAGCGGAGGAGISFSGNSNGTGGTGTG 1098	

RESULT 5  
B70807  
hypothetical glycine-rich protein Rv3512 - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: B70807  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: B70807  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1079 <COL>  
A:Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAAL17749.1; PID:g292444  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: Rv3512  
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology?

Query Match 35.1%; Score 1242; DB 2; Length 1079;  
Best Local Similarity 40.7%; Pred. No. 6.1e-63;  
Matches 316; Conservative 34; Mismatches 306; Indels 120; Gaps 29;

QY 5 TGGQMGGRGSMASGRGLGGCGAGAAAAAAGCA-GGGYGG-----LGSQG 53  
DB 58 TGGQMGNGN--GGNGTGGGTGGDGGALAGSSGAGGKGGNGDAGKAGTGSAPGTAG 115

QY 54 TSGRGLGGQ-----AGAAAAAAGCA-GGGYGG-----LGSQG 101  
DB 116 TGGDGGKGGGIGAGTTGPTGASGCTGSGGAGGTGGDGAANGGTAGAGGAGNG 175

QY 102 -----GAGAAAAAAGGAGGGYGG-----LGSQTSRGLGGCGAGAAA 145  
DB 176 KGGDGGAGVTSSTAGNSGAGGSGKGGACAGGAGATPGANGIAGNGDGD--GAAG 233

QY 146 AAAAAAGAGCGYGLGSG-----GTSGPGYGPQQTSRGGGLGGCGAGAAAAA 201  
DB 234 AVGISGATGAGDGGHGTGNAAGNGGTGGAGGSGIDYGGGTGTGGNGNGAIGCAGD 293

QY 202 AGGAG-CCGYGGLGSG-----GTSGRGLGGQ--AGAAAAAAGAGAG 244  
DB 294 AGGSNSGNGGIGKGNAGAGGAGNGTGVANGTGGDGGNGAAGATAGSNGGAG 353

QY 245 QGYGGLGSQTSRGLGGQ-----AGAAAAAAGAGAG-----QGYG 286  
DB 354 TGSAGNG--GTGGRGSGGAGDGGIGVGGKGNAGADGVEVGGAGGAGGSPNTPSGN 411

QY 287 GGLGSGQTSRGLGGCGAGAAAAAAGAGCGYGLGSGQTSQP---GGYGPQQ 343  
DB 412 GGQGGGSGGAG-GAAGAGGAGGANGTAGNGGQGGAGGTGGAGAASSATNGSGGAGG 470

QY 344 TSGRGLGGCGAGAAAAAAGAGCGYGLGSGQTSRGLGGQ-----A 393  
DB 471 TGGDGGCGAGTGGAGTGGAGDGGCGGQGGAGG--GAGGQGGAGGAGGTGGNGNITG 529

QY 394 GAAAAAAGAG--AGQYGGGLGSG--GTSGRGLGGCGAGAAAAAAGAGCGY 450  
DB 530 GTAGTAGAAGNGAAGKGGAGCGGTGGTGGQGGAGGAGGAGGTGGRTVGGGTVPAGS 589

QY 451 GGLGSGQTSRGLGGCGAGAAAAAAGAGCGYGLG-----SQGTSPPGCGP 505  
DB 590 GGQ--GNAGGGGAGG--GGADGSGDGGDAGTGGNGNGNRNSNGTGGAGNGGG 645

QY 506 QTSRGLGGCGAGAAA-AAAAAAGAGCGYGG--GLGSGQTSRGLGGCGAGAAAA 563  
DB 646 GAGGAGAGGSGGTGGNGCAGCGDAGDAGNGNGNGTGGNGNGGAGGAGT 705

QY 564 AAAAAAGAGCGYGL-----GSQGT--SGRGLGGCGAGAAAAAAGAGCGYGLG 618  
DB 706 GSGNGGSGGNGGAGCGGNSGTGSGDGGAGNGGAAGTGGTGGDGLGTGTGTGSGG 765

QY 619 SQGTSRGLGGCGAGAAAAAAGAGCGYGLGSGQTSRGGYGPQQTSG 674

DB 766 --GTGDCGNGGNGADNTANMTAAGDGGNGDGGFGGAGAGGGGLTAGANGTG 819

RESULT 6  
A70934  
hypothetical glycine-rich protein Rv0578c - Mycobacterium tuberculosis (strain H37Rv)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: A70934  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: A70934  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1306 <COL>  
A:Cross-references: GB:AL021942; GB:AL123456; NID:g3242298; PIDN:CAAL17449.1; PID:g290  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: Rv0578c  
C:Superfamily: collagen alpha 1(IV) chain

Query Match 33.8%; Score 1198.5; DB 2; Length 1306;  
Best Local Similarity 38.4%; Pred. No. 1.9e-60;  
Matches 323; Conservative 37; Mismatches 293; Indels 189; Gaps 34;

QY 11 GRGSMASGRGGLGGCGAGAAAAAAGG-----AGQGYGGLGSGQTSRGLGGCGAG 66  
DB 399 GTTTSGGNGDGGKADAISSGQTGANGRGDGGQVNGGAG--GAGRGGAGGLGFG 456

QY 67 AAAAAAAGG-AGQGYGG--LGSQTSRGLGGQGA-----GAAAAAAGAAA 114  
DB 457 SEAPRPGCAGGTGAGGNGGTQAGDGTGGAGGAGDGGSGGAGSICFNASAPCAAGSP 516

QY 115 GG-----AGQGYGGL-----GSQTSRGLGGQGA-----GAAA--- 146  
DB 517 GGNNGNGPGGAGGEGGAGGLAALASQNGSQGAGDGGAGNGGTPNGHGHAAGALGV 576

QY 147 --AAAAAGGAGCGYGLGSGQTSQ--PGYGP-PCQ--QTSRGLGGCGAGAAAAA 199  
DB 577 NGYGGAGGHHGDPGVGGAGGCGSSSTPGANGAPGNTPTSGNGNGNGRGADATGFGQT 636

QY 200 AAAG-----AGQGYGGLGSGQ-----GTSGRGLGGCGAGAAA 234  
DB 637 GASGGRGDDGGLVGNAGGAGGNGSKGLPLGLRLGNPLDGTGCGNGAGSGGAWAGN 696

QY 235 AAAAAAGG-AGQGYGGLGSGQ-----GTSGRGLGGQ--AGAAAAAAGAAA 277  
DB 697 GTTGAGGTGGVGTGGSGDGVNSSAGADGHPGCTGGVGTGGKGGDGGDGAAPNGV 756

QY 278 AGGAGCGYGLGSGQTSRGLGGCGAGAAAAAAG--CAGCGYCGGLGSG--GTSGP 335  
DB 757 AGSQPGCAGGAGD--GTGVGNGGCGIDGADGATAGAGDGGAGGAGGCGGTGTP 814

QY 336 GYGP-----GQTSRGLGGQ-----AGAAAAAAGAAA 364  
DB 815 GCAGPAGTTGSGAGGNGSGTGGDGDGNGANGSVFTNNGIGGNGNGNAGPSCAG 874

QY 365 AAGGA-----GQGYGLGSGQTSRGLGGCGAGAAAAAAGAAA 407  
DB 875 GSGAGSTFGATGSSSSITHVNGNGNGNGNDHALUSGNAAGNGNGNGNSLRSGGAG 934

QY 408 GQGYGGLGSGQTSRGLGGCGAGAAAAAAGAGAGCGYGG-----LGSQTS 460  
DB 935 GHGNGGNASRGMDGDTGGAGGNAGQINGGAGNGDGGTGSNGPALTGSGRGG 994

QY 461 RGLGGCGAGAAAAAAGAGCGYGLGSGQTS-----SGPGYGPQQTS 509

```
Db 995 DGVGGGGSVAGDADGGRGGA--GCTGTGLRGTGTTGATGATGTTFDAGADHGNGGTG 1052
QY 510 GRGGLGGGAGAAAAA-----GGAGGGYGGGLGSGQTSRGRGLGGGGA-G 558
Db 1053 GVGTTGAGGGGNGGAGGKALSPTGNNSSQAGAGDGGAGGAG--GTGTTGGDGRGAHG 1110
QY 559 AAAAAAAGAGAGGGYGGGLGSGQTSRGRGLGGGAGAAAAAAGAGAGGGYGGGLG 618
Db 1111 TLFSSLAGTGGTGGNGTGGTG--GTGGAGGAGGTGTTGATGATGAAGRAGNGVGGSG 1168
QY 619 SQTSS--GRGLGGGAGAAAAAAGAGAGGGYGGGLG-----SQTSPGGY-GPQQQT 672
Db 1169 GLGSATGPGTGGMG--GAGGTSTVSAGDGRGFGGDLADSSGNGGDDGGHGGDGFRT 1227
QY 673 SG 674
Db 1228 AG 1229

RESULT 7
E70806
hypothetical glycine-rich protein Rv3507 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: E70806
R: Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J. E.; Taylor, K.; Whitehead, S.; Barrell, B. G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: E70806
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-1381 <COL>
A: Cross-references: GB:AL020222; GB:AL123456; NID: g3261554; PIDN: CAAL1744.1; PID: g292444
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: Rv3507
C: Superfamily: collagen alpha 1(IV) chain

Query Match 33.8%; Score 1196.5; DB 2; Length 1381;
Best Local Similarity 38.7%; Pred. No. 2.5e-60;
Matches 319; Conservative 36; Mismatches 276; Indels 193; Gaps 36;

QY 7 GQOMGRGSMAS--GRGGLGGGAGAAAAAAGG-----AGQGGYGGGLGSGQ-----GTS 55
Db 150 GQSGGNGGSAGLWNGGAGGAGGSGGGAAGGNGCGWLFAGAGTGGTGTGAPGAMGGTG 209
QY 56 GRGG-----LGGGAGAAAAAAGAGAGAGGGYGG-----LGSQTSRGRGLGGGAGAG 105
Db 210 GNGNGGALLIGGGGLGAGAGMGGTGGTGGTGGNGNGALLIGAGGVGGAGGIGGQGTGA 269
QY 106 AAAAAAAGAGAGAGGGYGGGLGSGQTSRGRGLGGGAGAAAAAAGG-----AGGGYGG 161
Db 270 GGNAGAGGTG--GNGGAGGLFMNG--GDSGAGGCGGAGGAGGAGGAGGAGGAGGAGG 325
QY 162 GLGSQTSPPGGYGPQQQTSGRGLGGGAGAAAAAAGAGAGGGYGGGLGSGQTS--- 218
Db 326 GTGAGGAGPVLFGH-----GAGGMMGQG-----GTGGMGAGGDDCTTVI 366
QY 219 -SRGRGLGGGAGAAAAAAGAGAGAGGGY--GGGLGSGQTSRGRGLGGGAGAAAAAAG 276
Db 367 AGTGGEGGTGGAAGAGGAGAGGAGTSGGLAGGVGAGGTGGTGGTGGGAGGAGGAGGAG 426
QY 277 AAG-----GAGGGYGGGLGSGQ-----GTSRGRGLGSGQ----- 303
Db 427 ANGDPFAGKGGNGGIGGAAYTGGVAGDGGTGGKGTGAGGAGGAGNDAGSTGNPCKGGD 486
QY 304 ----GAGAAAAAAGG-----AGGGYGGGLGSGQTSG-----PGGYG- 339
Db 487 GGIGGAGGAGGAGTCTGCGHAGTGTGGDGGTGGNGGTGGVNGADNTLNPTTPGAGE 546
```

```
QY 340 PCQQTSGRGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 399
Db 547 PG-----GAGGAG--GAGGAAGPGGTGGTGGNGGNGNG--GNGGNGGNGGNGAGNNS 598
QY 400 AAAAAAG-----AGGGYGG-----GSGQTSRGRGLGGGAGAAAAAAGAGAGAGAG 449
Db 599 TNAVPGEGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 648
QY 450 YGGLGSGQTSG-----RGGLGGGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 498
Db 649 KACTGNSGNFGVDEAGFSGGAGNGGVGAAGAGGTGCGSGGNGGPDGAGGAGGAGGAG 708
QY 499 -PG-GYGPQQQTSGRGLGGGAGAGAAAAAAGG-----AGGGYGGGLGSGQ----- 544
Db 709 IPTGTGTEPAGGTGAKGGDGGDGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 768
QY 545 -----GTSRGRGLGGGAG-----GAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 584
Db 769 GKAPHGDAGSGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 828
QY 585 S-----GRGGLGGGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 640
Db 829 TITVPGNNGNAGDGN---CGNAGAGGNGGSGDFGNGNTTSGAGSGGNGG--NAGTAGSGG 884
QY 641 AAAAAAGAG-AGGYGGGLGSGQTSPPGGYGPQQQTSGTIRIRAPSTYS 683
Db 885 AGGTGTGLSGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNG 927

RESULT 8
A70869
hypothetical glycine-rich protein Rv2490c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: A70869
R: Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J. E.; Taylor, K.; Whitehead, S.; Barrell, B. G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: A70869
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-1660 <COL>
A: Cross-references: GB:AL021246; GB:AL123456; NID: g3261507; PIDN: CAAL16067.1; PID: g279
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: Rv2490c
C: Superfamily: collagen alpha 1(IV) chain

Query Match 33.3%; Score 1180; DB 2; Length 1660;
Best Local Similarity 37.6%; Pred. No. 2.4e-59;
Matches 326; Conservative 38; Mismatches 292; Indels 212; Gaps 35;

QY 2 ASMTGQQMGRGSMASGRGLGGGAGAGAAAAAAGAGAG-----GAG 45
Db 514 SSCTPGEDGNGG-AGGAGGAGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 572
QY 46 YGGLGSGQTSRGRGLGGGAGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 104
Db 573 NGGDG--GAGGRGGDGGAG--GAGGDAPAGRAGSQGVGGDGGAGGAGGAGGAGGAGGAG 629
QY 105 A-----AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 147
Db 630 AFKDDGGAGCGDGDGPFAGGKGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 689
QY 148 AAAAAAG-----GQGGYGGGLGSGQTSPPGGYGPQQ-----TSGRG 184
Db 690 ANGKGGAGGNGGLVGGDGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 749
```



A:Accession: E70917  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1329 <COL>  
A:Cross-references: GB:295844; GB:AL123456; NID:g3250713; PIDN:CAB09271.1; PID:g2131046  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: Rv1450c  
C:Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology

```
Query Match      32.5%; Score 1152; DB 2; Length 1329;
Best Local Similarity 35.8%; Pred. No. 7.6e-58;
Matches 326; Conservative 35; Mismatches 305; Indels 244; Gaps 35;

QY  7  GQMGRCMSASGRGLGGGQA-----GAAAAAAAAGGA-----  41
Db 133  GQAGGAGIILWNGGAGSGAPGVGGAGGAAGLFTGTGAGGAGGAGGAGGSGGWL 192
QY  42  GGGYGGGLGSGQTSRGLGGGAGAAAAAAGGAGGGYGGGLGSGQTSRGLGGQ 101
Db 193  GNGVGGAGGQSLLG-CATGAGGNAGLFGVGTGGPGPGPGVG--GTGGAGGLGT 249
QY 102  GAGAAAAAAAAGG-AGGGYGGGLGSGQ-----GTSRGLGGGAGAAAAA 148
Db 250  LYAGGHHGAGGPGPIGGVGGHGGVGAAGLLGVGGHGGAGGAGCAEVAGAAGEDLSPHG 309
QY 149  --AAAGCAGCGYGGGLGS--QGTSPGGYGPQGTSGRGLG-----GQAGAAAA 196
Db 310  TSGVGGDAGDGTGGRGGLAGAGGAGGAGGCGGTGAGGAGFSRALIVAGDNGDPGA 369
QY 197  AAAAAAGGAGQ--GGYGGGLGSGQTS-CRGLGGGAGAAAAAAGGA-----GQ 246
Db 370  GGAGCTGGAGSTGAHGAAGASPTSGNGGAGGAGNAHFSGGKAGGNGGAGGGLVNG 429
QY 247  YGGLGSGQGT-----SRGLGGGAGAAAAAAGGA-CGGYGGGLGS 291
Db 430  GAGGAGNGAPAPPSSGDPNNGGGAGGAGGKGDDGAQAGDGGAGGKGGNGGA 489
QY 292  QGTSRGL-----GGGAGAAAAAAGGA-----AGGA-----  319
Db 490  TGATGLNGLAGADGTDGKGGNGAGGGGGGAGGQKALAATHQDGMGAGGAGNGGA 549
QY 320  -GGYGGGLGSGQGT-----SGPGYGPQGTSGRGLGGGGA-----  355
Db 550  GMGGCGGCAKGTFTDNGDVGNGNGNGSGRTGGAGGIGGAGTAGAGCARGATP 609
QY 356  -----GAAAAAAAAGCAGCGYGLL-GSQGTSRGLGGGAGAAAAAAGG 406
Db 610  GNGGTGGGANATVAGGAGGAGGKGGNGLVNGGAGGKGGDMGVAGSSPTTAGESGT 669
QY 407  AGQ-----GGYGGGLGSGQ-----GTSRGLGGGAGAAAAAAGG-----  444
Db 670  SGNNGAGGAGGAGGGRGDFGGDGTGGAGNGANGANGATTPCAKGDGHHGPGAQGN 729
QY 445  AGGGYGGGL-----GSGQTSRGLGGGAGAAAAAAGGAGGAGGAGY- 489
Db 730  GGGGPGGLAGNLFPQNGIQCVGSGKKGAGGLAGDGGNANGCNFPFGDNGHGGNG 789
QY 490  --GLGSGQTSRGGYGPQ-----TSRGLGGGAGAAAAAAGGAGGCGY 539
Db 790  NPGAGGGGSGGAGSTPCAAGHGTFTPSGDDGDDGNGGN-----SQVYVNGGDDGNG 844
QY 540  GL-GSQGTSRGLGGGAGAAAAAAGG-----AAAGCAGCGYGGGLGSGQ 583
Db 845  GNGSAGCTGNGGGRGGDGFAGGMSANATNPNCPNGPNPNPNGCAGGAGGAGLNG-GNGG 903
QY 584  TSGRGLGGGAGAAAAAAGGAGQ-----GGYGGGL-----SQGTSGR 625
Db 904  AGNNGGLGFGGAGCAANGVAVAPQPGGAGGHHGAGGNGGAGGNGGVVSDGAGGA 963
QY 626  GGLGGG-----AGAAAAAAAAGCAGCGYGG--LGSQGTSGPGYCPGCGTSGIR 676
Db 964  GGAGDGGAPDGNANGNGGAGAFAGGGGGRGDDGNAGNAGGPGGTGTAGKAG-- 1021
```

QY 677 IRAPSTFEH 686  
Db 1022 ---PAGSILH 1028

# RESULT 11

A70896  
hypothetical glycine-rich protein Rv1091 - Mycobacterium tuberculosis (strain H37Rv)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 24-Nov-1999  
C:Accession: A70896  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: A70896  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-853 <COL>  
A:Cross-references: GB:AL021897; GB:AL123456; NID:g3256022; PIDN:CAA17207.1; PID:e125  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: Rv1091  
C:Superfamily: unassigned collagens

```
Query Match      32.5%; Score 1151; DB 2; Length 853;
Best Local Similarity 40.4%; Pred. No. 6.3e-58;
Matches 301; Conservative 27; Mismatches 307; Indels 110; Gaps 28;

QY  7  GQMGRCMSASGRGLGGGAGAAAA-----AAAAAGCAGCGYGGGLGSGQTS-----  55
Db 131  GQDGGPGLLYGNGNGTSTTAGVAGNGCAAGLNGGAGGGGAGAGGAGGAGGWL 190
QY  56  -GRGLGGGAGAAAAAAGGA-----GOGYGGGLGSGQTS-----RGLGGGGA 103
Db 191  YNGGAGGAGGTSVIPVAGNGGAGGAGSAGLWGTGAGDGGNGRSGPVPVAGSAGNGG 250
QY 104  GAAAAAAAAGGAGCGYGGGLG-----SGQTSRGLGGGAGAAAAAAGGA 155
Db 251  AGGAAGLFGDAGAGNGKGGAGAAFINFTAGDGGAGGAGGGHALLMGAGGAGNG 310
QY 156  GGGYGGGL--SQGTSPPGYGPQQTs-----GRGLGGGA-----GAAAAAAAAGGA 205
Db 311  GSGGTGAGGAGTACAGCGNGGAGGGGTGGLLFGNGAGGAGHCAAGNGLAAGNGVSSGGG 370
QY 206  GGGYGGGLGSGQTSRGLGGG-----AGAAAAAAGGAGCGYGGGL-----GSQ 254
Db 371  GAGGTGGAGDGG--GAGGAGNARLWVGAGAGGAGDGGAGGAGCGGSLSGNANGAG 428
QY 255  GTSRGLGGGAGAAAAAAGGAGCGYGGGLGSGQTSRGLGGGAGAAAAAAGGA 314
Db 429  GDSRGGTG--GAGGEGAGLLVGTGHHGDDGAG--GAAYKGGDGAAGATGTAGAGG 484
QY 315  AAGCAGCGYGGGLGSGQTSPPGYGPQQTSGRGLGGGAGAAAAAAGG-----  368
Db 485  RGGAGGGSGGDDGGGAAGPAGW-----LFGDGGAGNGGAGGAGGAGGAGGNGG 539
QY 369  -AGGGYGGGLGSGQTS-----GRGLGGGAGAAAAAAGGA-----AAGCAGCGYGGGLGSG 418
Db 540  NGNGNGNGNGNGATCGWLNGGAGCGCATAGAGAGAGNGVSTNGGTGCGNGGIG--  597
QY 419  GTSRGLGGGAG--GAAAAAAAAGGAG--OGYGGGL-----SQGTSRGLGGG--CAG 470
Db 598  GTGSGGAGGAGNAGLLVGGAGGAGGAGDGRGAGGTGFTISSDGGAGGDDGGNGNG 657
QY 471  AAAAAAAAAGGAGCGYGGGLGSGQTSPPGYGPQQTSGRGLGGGAGAAAAAAGGA 530
Db 658  GTGGLLFGAGNGGPGGSGGNAIDIG--GNGAGNGGTGDNNGNGSGGG--AGSGDGG 712
```









GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:06:04 : Search time 10.6063 Seconds  
(without alignments)  
2702.183 Million cell updates/sec

Title: US-09-490-291-6

Perfect score: 3543

Sequence: 1 MASMTGGQMGGRSMASGRG.....TSGIRAPSTFEHHHHH 691

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2138	60.3	747	1 SPDL_NEPCL	P19837 nephila cla
2	1281.5	36.2	1901	1 Y208_MYCTU	O53553 mycobacteri
3	1174.5	33.1	5263	1 FBOH_BOMMO	P05790 bombyx mori
4	1138.5	32.1	627	1 SPDL_NEPCL	P46804 nephila cla
5	1110.5	31.3	914	1 WA22_MYCTU	O06794 mycobacteri
6	1049.5	29.6	778	1 Y234_MYCTU	P71933 mycobacteri
7	1032	29.1	957	1 Y278_MYCTU	P56877 mycobacteri
8	988.5	27.9	801	1 Y747_MYCTU	O53810 mycobacteri
9	976	27.5	860	1 ELS_MOUSE	P54320 mus musculu
10	939	26.5	864	1 ELS_RAT	O99372 rattus norv
11	869	24.5	603	1 YD25_MYCTU	Q10637 mycobacteri
12	853.5	24.1	465	1 GRP2_PRAVU	P10496 phaseolus v
13	777.5	21.9	747	1 ELS_BOVIN	P04985 bos taurus
14	763	21.5	641	1 EBN1_EBV	P03211 epstein-bar
15	747.5	21.1	1356	1 CA21_ONCMY	O93484 oncorhynch
16	709	20.0	750	1 ELS_CHICK	P07916 gallus gall
17	693.5	19.6	1355	1 CA21_RANCA	O42350 fana catesb
18	691	19.5	730	1 ELS_HUMAN	P15502 homo sapien
19	686	19.4	498	1 Y118_MYCTU	O50615 mycobacteri
20	681	19.2	1733	1 VNVA_PRVKA	P33485 pseudorabie
21	678.5	19.2	543	1 YP91_PETHY	O50630 mycobacteri
22	673	19.0	384	1 GRP1_PETHY	O90789 petunia hyb
23	669	18.9	481	1 LORI_MOUSE	P18165 mus musculu
24	656.5	18.5	1027	1 CAFF_RIFPA	P30754 riftia pach
25	644	18.2	338	1 GRP_ARATH	P27483 arabidopsis
26	630	17.8	1372	1 CA21_MOUSE	Q01149 mus musculu
27	626	17.7	491	1 YK98_MYCTU	Q10707 mycobacteri
28	626	17.7	515	1 Y140_MYCTU	O50594 mycobacteri
29	624.5	17.6	1262	1 CA13_CHICK	P12105 gallus gall
30	622	17.6	1466	1 CA13_HUMAN	P02461 homo sapien
31	608	17.2	1372	1 CA21_RAT	P02466 rattus norv
32	608	17.2	1418	1 CA12_HUMAN	P02458 homo sapien
33	607	17.1	1464	1 CA13_MOUSE	P08121 mus musculu

#### ALIGNMENTS

##### RESULT 1

ID	SPDL_NEPCL	STANDARD;	PRT;	747 AA.
AC	P19837;			
DT	01-FEB-1991 (Rel. 17, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Spidroin 1 (Dragline silk fibroin 1) (Fragment).			
OS	Nephila clavipes (Orb spider).			
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;			
OC	Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.			
OX	NCBI_TaxID=6915;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE=90384959; PubMed=2402494;			
RA	Xu M., Lewis R.V.;			
RT	'Structure of a protein superfiber: spider dragline silk.';			
RL	Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).			
RN	[2]			
RP	SEQUENCE OF 653-747 FROM N.A.			
RX	MEDLINE=94165038; PubMed=8120021;			
RA	Beckwith R., Arcidiacono S.;			
RT	'Sequence conservation in the C-terminal region of spider silk proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and Araneus bicentenarius (Araneidae).';			
RL	J. Biol. Chem. 269:6661-6663(1994).			
CC	-1- FUNCTION: Spiders major ampullate silk possesses unique characteristics of strength and elasticity. Fibroin consists of pseudocrystalline regions of antiparallel beta-sheet interspersed with elastic amorphous segments.			
CC	-1- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 2, OF THE DRAGLINE SILK.			
CC	-1- DOMAIN: Highly repetitive protein characterized by regions of polyalanine and glycine-rich repeating units.			
CC	-1- SIMILARITY: BELONGS TO THE SILK FIBROIN FAMILY.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	EMBL; M37137; AAA29380.1; -			
DR	EMBL; U03848; AAB60212.1; -			
DR	PIR; A36068; A36068.			
KW	Silk; Repeat.			
FT	NON_TER	1	1	25 X APPROXIMATE TANDEM REPEATS.
FT	DOMAIN	1	655	1.
FT	REPEAT	1	25	2.
FT	REPEAT	26	38	3.
FT	REPEAT	39	66	4.
FT	REPEAT	67	96	5.
FT	REPEAT	97	130	6.
FT	REPEAT	131	158	

P02457 gallus gall  
P04258 bos taurus  
O53416 mycobacteri  
Q9xs17 canis famil  
P08123 homo sapien  
P28481 mus musculu  
P02452 homo sapien  
O46332 canis famil  
P02465 bos taurus  
P11087 mus musculu  
P02467 gallus gall  
P17140 caenorhabdi

34 606.5 17.1 1453 1 CA11\_CHICK  
35 603.5 17.0 1049 1 CA13\_BOVIN  
36 602.5 17.0 483 1 YA68\_MYCTU  
37 602.5 17.0 1460 1 CA11\_CANFA  
38 602 17.0 1366 1 CA21\_HUMAN  
39 599.5 16.9 1459 1 CA12\_MOUSE  
40 599.5 16.9 1464 1 CA11\_HUMAN  
41 599 16.9 1366 1 CA21\_CANFA  
42 595 16.8 1384 1 CA21\_BOVIN  
43 592.5 16.7 1453 1 CA11\_MOUSE  
44 573.5 16.2 1362 1 CA21\_CHICK  
45 563 15.9 1758 1 CA24\_CAEEL

```
FT REPEAT 159 191 7.
FT REPEAT 192 204 8.
FT REPEAT 205 235 9.
FT REPEAT 236 262 10.
FT REPEAT 263 292 11.
FT REPEAT 293 305 12.
FT REPEAT 306 333 13.
FT REPEAT 334 360 14.
FT REPEAT 361 394 15.
FT REPEAT 395 424 16.
FT REPEAT 425 458 17.
FT REPEAT 459 485 18.
FT REPEAT 486 512 19.
FT REPEAT 513 525 20.
FT REPEAT 526 555 21.
FT REPEAT 556 582 22.
FT REPEAT 583 612 23.
FT REPEAT 613 642 24.
FT REPEAT 643 655 25.
FT CONFLICT 662 662
FT CONFLICT 672 672
FT CONFLICT 695 747
FT
FT
SQ SEQUENCE 747 AA; 60528 MW; 850E44B0D649E012 CRC64;

Query Match
Best Local Similarity 60.3%; Score 2138; DB 1; Length 747;
Matches 498; Conservative 20; Mismatches 49; Indels 184; Gaps 35;

QY 25 QGAGAAAAAAGAGAGAGGGYGLGSGQT--SGRGLGGGAG-AAAAAAGAGGAGQ 81
D 1 QGAG-----AAAAAGAGAGGGYGLGGGAGGGYGLGGGAGGAGAGAAAAAGAGQ 56
QY 82 GYGGLGSGTSGRGLGGGAGAAAAAAGAGAGGGYGLGSGTSGRGLGGGAG 141
D 57 GYGGLGSGG-----AGRGQGAG-----AAAAAGAGGGYGLGSGG-AGRGLGGGQ 107
QY 142 GAAAAAAGAGAGGGYGLGSGTSGPGYGPQQTSGRGLGGGAGAAAAAAG 201
D 108 G-----AAAAAGAGGGYGLGNQG-----AGRGQG-----AAAA 142
QY 202 AGAGGGYGLGSGTSGRGLGGGAGAAAAAAG-----AGAGGGYGLG 252
D 143 AGAGGGYGLGSGG-AGRGLGGGAGAAAAAAGAGAGGGYGLGGGAGGGYGLG 201
QY 253 SGTSGRGLGGGAGAAAAA-----AAAAAGAGGGYGLG 291
D 202 SQG-AGRGLGGGAGAAAAAAGAGAGGGYGLGGGAGAGAAAAAGAGGGYGLG 260
QY 292 QTSRGLGGGAGAAAAAAGAGAGGGYGLGSGTSGPGYGP-GQTSRGLG 350
D 261 QG-----AGRGEGAG-----AAAAAGAGGGYGLGGG-AGQGYGLGSGAGRGL 311
QY 351 GCGAGAAAA-----AAAAAGAGGGYGLGSGTSGRGLGGG 392
D 312 GCGAGAAAAAGAGGGYGLGGGAGAGAAAAAGAGGGYGLGSG-AGRGLGGG 370
QY 393 AGAAAAAAGAGAGGGYGLGSGTSGRGLGGGAGAAAAAAGAGAGGGYGG 452
D 371 AG---AAGAGAGGGYGLGSGG-----AGRGQGAG-----AAAAAGAGGGYGG 419
QY 453 LGSQTSRGLGGGAGAAAAAAGAGAGGGYGLGSGT-----SGP 499
D 420 LGNQG-AGRGLGGGAG---AAAAAAGAGGGYGLGNQAGRGGGAGAAAAAGGAGQ 475
QY 500 GYGP-GQTSRGLGGGAGAAAAAAGAGAGGGYGLGSGTSGRGLG 552
D 476 GYGGLGSGAGR---GGGAGAGAAAAAAGAGAGGGYGLGSGG-AGRGLG 531
QY 553 GCGAGAAAA-----AAAAAGAGGGYGLGSGTSGRGLG 592
D 532 GCGAGAAAAAGAGGGYGLGGGAGAGAAAAAAGAGGGYGLGSGG-----AGRG 587
```

```
QY 593 QGAGAAAAAAGAGAGGGYGLGSGTSGRGLGGGAGAAAAAAGAGAGGGY 652
D 588 QGAG-----AAAAAGAGAGGGYGLGGGAG-GRGLGGGAG-----AAAGAGGGY 636
QY 653 GGLGSGTSGPGYGPQQTSGIRIRAPSTS 683
D 637 GGVGGSAA-----SAAASRLSSPQAS 659

RESULT 2
YZ08_MYCTU
ID YZ08_MYCTU STANDARD; PRT; 1901 AA.
AC O53553;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PE-PGRS family protein RV3508 precursor.
GN RV3508 OR MTU023.15.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basmah D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL022022; CAAL17745.1; -.
DR TubercuList; RV3508; -.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE_region; 1.
KW Hypothetical protein; Repeat; Signal; Complete proteome.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1901 HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT RV3508.
SQ SEQUENCE 1901 AA; 147627 MW; C7B1923D5D0146CD CRC64;

Query Match 36.2%; Score 1281.5; DB 1; Length 1901;
Best Local Similarity 39.5%; Pred. NO. 6.5e-51;
Matches 354; Conservative 28; Mismatches 273; Indels 241; Gaps 39;

QY 7 GOOMGRGSNASRGGGLGGQ-----GAGAAAAAAGAGAGGGYGLGSGQ----- 52
D 280 GCGGAGGAGSGGAGLGGTGGTGGAGGAGGALLLGGGGLGGGAGGAGGAGGAG 339
QY 53 -----GTSRGLGG-----QGAGAAA-----AAAAAAGG-AGCGGGYGLGSG 90
D 340 GGVVGGYGGTGGGGVAGLGGAGGAGGAGLFGNAGGAGAGVGGTGGGAGGAGGAG 399
QY 91 G-----TSGRGLGGGAGAAAAAAGAGAGGGYGLGSGQTSR----- 133
```

```
Db 400 GADAPASTGLTGTGTFAGGAGGCGGGNAIAGGINSGGAGGTGGGGAGGGSGADN 459
QY 134 -----GGLGGGQ--AGAAAAAAAAGG-----AGCGGYGGLGSGQT 168
Db 460 ASGTGADGAGGAGTGTGGVGAAGRAGTGGTGGCGGAGGAGSAGT 519
QY 169 -----SGPGYGPQOOTSRRGLGGGAGAAAAAAGAGGAGGQ---- 207
Db 520 DATATGATGTGTSAGGAGGAGGNTGVTGNGSGGQ--GTGAGGAGGAGGADNPT 578
QY 208 --GGYGGGSGQTSRRGLGGQ--AGAAAAAAAAGG--AGCGGYGGLGSGQT---SGR 259
Db 579 GIGTGTGCKG--GAGGAGGSGSAGGTNGSGAGGTGGGAGGAGGADNPTGI 636
QY 260 GGLGGGAGAAAAAAGAGGAGGCGG--LGSQTSRRGLGGQ--GAGAAAAA 311
Db 637 GGAGGTGTTGAAGAGGAGGAGTGTGGVAVSVNAGIGTGTGGVGGAGGAGAAAA 696
QY 312 AAAAAGGA---GOGGYGGLGQ---GTSPPGYGPQOOTSRRGLGGQ--- 355
Db 697 GSSATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 756
QY 356 -----GAAAAAAGAGGA--GOGGYGGLGQ---GTSPPGYGPQOOTSRRGLGGQ--- 393
Db 757 AGGAGGTGGAAGAGGAGGAGTGTGGVAVSVNAGIGTGTGGVGGAGGAGGAGG 816
QY 394 -----GAAAAAAGAGGA--GOGGYGGLGSGQTSRRGLGGQ--GAGAAAA- 435
Db 817 FDGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 876
QY 436 -----GAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 475
Db 877 NGNAGVGLTAKAGDGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 936
QY 476 AAAAAGGA---GOGGYGGLGSGQTSRRGLGGQTSRRGLGGQAGAAAAAAG 531
Db 937 SINANGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 993
QY 532 GAGCGYGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 580
Db 994 GGGNGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1053
QY 581 SGTGSG---RGGLGGQ--GAGAAAA--AAAAAAGGAGGAGGAGGAGGAGGAGG 621
Db 1054 ATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1113
QY 622 --TSRRGLGGQGA---GAAAAAAGAGGA---GOGGYGGLGSGQTSRRGLGGQ 667
Db 1114 NFGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1169

RESULT 3
FB0H_BOMMO STANDARD; PRT; 5263 AA.
ID FB0H_BOMMO
AC P05790; Q26379; Q17220;
DT 01-NOV-1988 (Rel. 09, Created)
DE 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fibroin heavy chain precursor (Fib-H) (H-fibroin).
GN FBH.
OS Bombyx mori (Silk moth).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC Diptera; Bombycoidea; Bombycidae; Bombyx.
OX NCBI_TaxID=7091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20330362; PubMed=10871375;
RA Zhou C.-Z., Confalonieri F., Medina N., Zivanovic Y., Esnault C.,
RT Yang T., Jacquet M., Janin J., Duguet M., Perasso R., Li Z.-G.;
RT "Fine organization of Bombyx mori fibroin heavy chain gene.";
RL Nucleic Acids Res. 28:2413-2419(2000).
```

```
RN [2]
RP SEQUENCE OF 1-168 FROM N.A.
RX MEDLINE=80045039; PubMed=498286;
RA Tsujimoto Y., Suzuki Y.;
RT "The DNA sequence of Bombyx mori fibroin gene including the 5'
RT flanking, mRNA coding, entire intervening and fibroin protein coding
RT regions";
RL Cell 18:591-600(1979).
RN [3]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=79211211; PubMed=455439;
RA Tsujimoto Y., Suzuki Y.;
RT "Structural analysis of the fibroin gene at the 5' end and its
RT surrounding regions";
RL Cell 16:425-436(1979).
RN [4]
RP PARTIAL SEQUENCE FROM N.A.
RX STRAIN=Kinshu X Showa;
RX MEDLINE=89094868; PubMed=3210244;
RA Mita K., Ichimura S., Zama M., James T.C.;
RT "Specific codon usage pattern and its implications on the secondary
RT structure of silk fibroin mRNA.";
RL J. Mol. Biol. 203:917-925(1988).
RN [5]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=94365842; PubMed=7916056;
RA Mita K., Ichimura S., James T.C.;
RT "Highly repetitive structure and its organization of the silk fibroin
RT gene.";
RL J. Mol. Evol. 38:583-592(1994).
RN [6]
RP SEQUENCE OF 5179-5263 FROM N.A., AND DISULFIDE BONDS.
RX STRAIN=J-139;
RX MEDLINE=99296390; PubMed=10366732;
RA Tanaka K., Kajiyama N., Ishikura K., Waga S., Kikuchi A., Ohtomo K.,
RA Takagi T., Mizuno S.;
RT "Determination of the site of disulfide linkage between heavy and
RT light chains of silk fibroin produced by Bombyx mori.";
RL Biochim. Biophys. Acta 1432:92-103(1999).
CC -I- FUNCTION: FORMS THE SILK FILAMENT; A STRONG, INEXTENSIBLE,
CC INSOLUBLE AND CHEMICALLY INERT FIBRE.
CC -I- SUBUNIT: FORMED OF TWO CHAINS: HEAVY AND LIGHT, THAT ARE LINKED BY
CC A DISULFIDE BOND. HEAVY-LIGHT CHAIN ASSEMBLY IS ESSENTIAL FOR THE
CC EFFICIENT INTRACELLULAR TRANSPORT AND SECRETION OF FIBROIN.
CC -I- TISSUE SPECIFICITY: PRODUCED EXCLUSIVELY IN THE POSTERIOR (PSG)
CC SECTION OF SILK GLANDS.
CC -I- DOMAIN: COMPOSED OF ANTIPARALLEL BETA SHEETS. THE STRANDS OF THE
CC BETA SHEETS RUN PARALLEL TO THE FIBER AXIS. LONG STRETCHES OF SILK
CC FIBROIN ARE COMPOSED OF MICROCRYSTALLINE ARRAYS OF (-GLY-SER-GLY-
CC ALA-GLY-ALA-N) INTERRUPTED BY REGIONS CONTAINING BULKIER RESIDUES.
CC THE FIBER IS COMPOSED OF MICROCRYSTALLINE ARRAYS ALTERNATING WITH
CC AMORPHOUS REGIONS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF226688; AAF76983.1; --
DR EMBL; V00094; CAA23432.1; --
DR EMBL; V00097; CAA23433.1; --
DR EMBL; S74439; AAB31861.1; --
DR EMBL; X13869; CAA32076.1; --
DR EMBL; M35378; AAB27839.1; --
DR EMBL; AB017362; BAA33147.1; --
DR PIR; S01844; S01844.
KW Silk; Signal; Repeat.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 5263 FIBROIN HEAVY CHAIN.
FT DOMAIN 149 5206 HIGHLY REPETITIVE.
```

FT	DISULFID	5244	5244	INTERCHAIN (WITH LIGHT CHAIN).	
FT	DISULFID	5260	5263		
FT	CONFLICT	10	10	C -> V (IN REF. 2).	
SQ	SEQUENCE	5263	AA; 391586 MW; 8EEL1D3A0A47440E CRC64;		
Query Match					
Best Local Similarity					33.1%; Score 1174.5; DB 1; Length 5263;
Matches					318; Conservative 49; Mismatches 305; Indels 131; Gaps 27;
QY	2	ASMTGGQQMGSG-MASGRGLGGOGAGAAAAAAGAGAGGGYGGGLGSGQTSGRGLG	60		
DB	4337	ACAGAGYAGAGSGAASGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG	4396		
QY	61	GQGGACAAAAAAGGACG- - - - -GYG- - - - -GLGSGQTSGRGLGGOGAGAAAAA	112		
DB	4397	SGAGSGAGAGSGAGAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAG	4456		
QY	113	AGGAGAGGGYGGGLGSGQTSGRGLGGOGAGAAAAAAGAGAGG- - - - -GYG- - - - -GLGSGQTSG	170		
DB	4457	SGSGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG	4516		
QY	171	PG- - - - -GYGPGQQTSG- - - - -RGGLG- - - - -GOGAGAAAAAAGGAGG	207		
DB	4517	ACAGAGAGAGTSSGFGPYVANGYSGYEWYSSDFGTGSGAGSGAGSGAGAGY	4576		
QY	208	G- - - - -GYG- - - - -GLGSGQTSGRGLGGOGAGAAAAAAGAGAGG- - - - -GYG- - - - -GLGSGQTSG	253		
DB	4577	GAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAG	4636		
QY	254	QGTSGRGLGGOGAGAAAAAAGGAGG- - - - -GYG- - - - -	287		
DB	4637	GAGAGSGAGAGSGAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAG	4696		
QY	288	- - - - -GLGSGQTSGRGLGGOGAGAAAAAAGAGAGGGYGGGLGSGQTSGPG- - - - -GYGP	340		
DB	4697	GSGAGAGSGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGYGA	4756		
QY	341	QGTSGRGLG- - - - -GOGAGAAAAAAGAGAGGGYGGGLGSGQTSGRGLGG	391		
DB	4757	GYGIGVAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAG	4816		
QY	392	GAGAAAAAAGAGAGGGYGGGLGSGQTSGR- - - - -GLG- - - - -GOGAGAAAAAAG	443		
DB	4817	GAGAGSGAGAGSGAGAGSGAGAGSGAGAGYAGYAGYAGYAGYAGYAGYAG	4876		
QY	444	GAGGGYGGGLGSGQTSGRGLGGOGAGAAAAAAGAGAGG- - - - -GYG- - - - -GL	491		
DB	4877	GAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGAGYAGYAGY	4936		
QY	492	GSQGTSGPG- - - - -GYGPQQTSGRGLGGOGAGAAAAAAGAGAGGGYGGGLGSGQTSG	548		
DB	4937	GAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAG	4995		
QY	549	RGLGGOGAGAAAAAAGAGAGG- - - - -GYG- - - - -GLGSGQTSGRGLGGOGAGAAAAA	606		
DB	4996	AGSGAGSGAGYAGAGAGYAGAGAGYAGAGAGYAGAGAGYAGAGAGYAGAG	5055		
QY	607	GGAGG- - - - -GYG- - - - -GLGSGQTSGRGLGGOGAGAAAAAAGAGAGG- - - - -GYG	653		
DB	5056	SGAGSGAGAGSGAGYAGAGAGSGAASGAGAGSGAGAGAGAGAGAGAGAG	5115		
QY	654	- - - - -GLGSGQTSGPG- - - - -GYGPQQTSG	674		
DB	5116	AGAGSGAASGAGAGAGTSSG	5138		
RESULT 4					
ID	SPD2_NEPCL	STANDARD;	PRT;	627	AA.
AC	P46804;				
DT	01-NOV-1995 (Rel. 32, Created)				
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				

DE	Spidroin 2 (Dragline silk fibroin 2) (Fragment).				
OS	Nephila clavipes (Orb spider).				
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;				
OC	Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.				
OX	NCBI_TaxID=6915;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=92406876; PubMed=1527052;				
RA	Hinman M.B., Lewis R.V.:				
RT	"Isolation of a clone encoding a second dragline silk fibroin.				
RL	Nephila clavipes dragline silk is a two-protein fiber.";				
CC	J. Biol. Chem. 267:19320-19324(1992).				
CC	-!- FUNCTION: Spiders major ampullate silk possesses unique				
CC	characteristics of strength and elasticity. Fibroin consists of				
CC	pseudocrystalline regions of antiparallel beta-sheet interspersed				
CC	with elastic amorphous segments.				
CC	-!- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 1, OF THE DRAGLINE SILK.				
CC	-!- SUBCELLULAR LOCATION: Extracellular.				
CC	-!- DOMAIN: Highly repetitive protein characterized by regions of				
CC	polyalanine and glycine-rich repeating units.				
CC	-!- SIMILARITY: BELONGS TO THE SILK FIBROIN FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; M92913; AAA29381.1; -				
KW	Silk; Repeat.				
FT	DOMAIN	1	530	15 APPROXIMATE TANDEM REPEATS.	
FT	REPEAT	1	36	1.	
FT	REPEAT	37	79	2.	
FT	REPEAT	80	121	3.	
FT	REPEAT	122	172	4.	
FT	REPEAT	173	213	5.	
FT	REPEAT	214	252	6.	
FT	REPEAT	253	283	7.	
FT	REPEAT	284	317	8.	
FT	REPEAT	318	359	9.	
FT	REPEAT	360	391	10.	
FT	REPEAT	392	428	11.	
FT	REPEAT	429	464	12.	
FT	REPEAT	465	488	13.	
FT	REPEAT	489	515	14.	
FT	REPEAT	516	530	15.	
SQ	SEQUENCE	627	AA; 54184 MW; CB9B63779B2C594B CRC64;		
Query Match					
Best Local Similarity					32.1%; Score 1138.5; DB 1; Length 627;
Matches					325; Conservative 36; Mismatches 190; Indels 141; Gaps 36;
QY	23	GQGGACAAAAAAGGAGGGYGGGLGSGQTSGRGLGGOGAGAAAAAAG- - - - -G	78		
DB	2	GGYGPQ- - - - -QGPGYGP-GQGGPSGPG- - - - -SAAAAAAGPGGCG	42		
QY	79	AGQGGYGGGLG-SQGTSGRGLGGOGAGAAAAAAGAGAGGGYGGGLGSG- - - - -GTSGR	133		
DB	43	PQQQPGGYPGQGGPGYGP-GQGGPSGPGSAAAAA-GSGQGGPGYCPQQPGGCG	101		
QY	134	GGLGGOGAGAAAAAAGGAGGGYGGGLGSGQTSGRGLGGOGAGAAAAAAG- - - - -	187		
DB	102	QQGGPSGPGSAAAAAAGSQQGGPGYGP-PQQGGPGYGPQGG- - - - -GPGYGPQGGP	158		
QY	188	-GQGGACAAAAAAGGAGGGYGGGLGSGQTSGR- - - - -GLGGOGAGAAAAAAGG	242		
DB	159	SGPGSAAAAAAGSGPQQGGPGYGP-GQGGPGYGPQGGPSGPGSAAAAAAGSGPG	217		
QY	243	AGQGGYGGGLGSGQTSGRGLG-GQ- - - - -GAGAAAAAAGGAGGGYGGGLGSGQTSGR	297		

```

Db 218 GQPGGYG-----PGQPGGYGPGQGLSGPGSAAAAAAGPGQPGGYGP-GQQPGSPG 272
QY 298 GGLGGGAGAAAAAAGAGAGAGCGGCGGGLGSGQTSRGGYGPQQTSGRGLGGGQAGA 357
Db 273 G-----SAAAAAAGAGPGGYG-----PGQPGGYGPGQ-----GPGSAGS 310
QY 358 AAAAAAAGAGAGCGGGLGSGQTSRGG-----GLGGGAGAAAAAAGAGAGAGAGAG 412
Db 311 AAAAAAGAGCGGGLGSGYGP-GQQPGGYGPGQPGGYGPGSAAAAAAGPGQPGGY 369
QY 413 GGLGSGQTSRGGGLGGGAGAAAAAAGAGAGAGCGGCGGGLGSGQTSR-----RGGLGGGAG 470
Db 370 GP-GQQPGSPG-----SASAAAAAAGAGPGGYGP-GQQPGGYAPQPGSPG 417
QY 471 AAAAAAAGAGAGCGGCGGGLGSGQTSRGGYGPQQTSGRGLGGGAGAAAAAAGAG 530
Db 418 SASAAAAAAGAGPGGYG-----PGQPGGYGPGQ-----GPGSAGSAAAAAAG 463
QY 531 GGAGGCGGGLGSGQTSRGGGLGGGAGAAAAAAGAGAGAGCGGCGGGLGSGQTSR 590
Db 464 --AGPGGYGP-AQPGSPG-----IAASAAAGPGGYG-----PAQGP 500
QY 591 GGAGGAGAAAAAAGAGAGCGGCGGGLGSGQTSRGGGLGGGAGAAAAAAGAG 647
Db 501 AGYGPSSAASA-----GAGSAGYGP-GSQAASAAASRLASPDSCARVASVNLVSGPT 555
QY 648 GQGGYGLGSGQTSRGGYGPQQTSGIRIRA 679
Db 556 SSAALSSVISNAVSVIGASNPGLSGCDVLIQA 587

RESULT 5
WAZ22_MYCTU
ID WAZ22_MYCTU STANDARD; PRT; 914 AA.
AC 006794;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE WAZ22 antigen precursor.
GN WAZ22 OR RV1759C OR W1807 OR WTCY28.25C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID:1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC SUBFAMILY.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A

```

```

CC FRAMESHIFT IN POSITION 85.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z55890; CAB09322.1; -.
DR EMBL; AE007040; -. NOT_ANNOTATED_CDS.
DR TIGR; MT1807; -.
DR TubercuList; Rv1759c; -.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE; 1.
KW Antigen; Repeat; Signal; Complete proteome.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 914 WAZ22 ANTIGEN
SQ SEQUENCE 914 AA; 74354 MW; F6953C3DBE8E6AC8 CRC64;

Query Match 31.3%; Score 1110.5; DB 1; Length 914;
Best Local Similarity 38.2%; Pred. No. 1.6e-43;
Matches 309; Conservative 31; Mismatches 307; Indels 161; Gaps 33;

QY 5 TGGQMGREGSWASRGGLGGGAGAAAAAAGAGAG-----QGGYGLG-SSQGTSGRG 59
Db 110 TGRPLGNG--ANGAPGTGANGDAGWLIGNGGAGGSCAKGANGGAGGPGGAAGLFGNGG 167
QY 60 LGGOGAGAAAAAAGAG-----GOGYGGGLGSGQTS---GRGLGGGAGAAAAA 109
Db 168 AGGAGGTATTANGIGGAGGAGGSAAMLFAGAGGAGGAGGAATSLVIGIGTGTGTGNAGMLA 227
QY 110 AAAAAGAG-----QGGYGGGLGSGQ-----TSRGGGLGGQ-----ACAAAAAAG 154
Db 228 GAAGAGAGGAGGFSFTAGGAGGAGGAGGLFTTGGVGGAGGQHTGGAGGAGGAGLFGAGG 287
QY 155 -AGOGYGGGLGSGQTSRGGYGPQQTSGRGLGGGAGAAAAAAGAG-----204
Db 288 MGGAGGFGDHTLTGTGGAGGDDGGGLFGAGGDDGAGGSGLTTCGGAAGNGNAGTSLGA 347
QY 205 -AGOGYGGGLGSGQ-----SRRGGGLGGGAGAAAAAAGAGAG-----OGYGLG 253
Db 348 AGGAGGTGGAG--GTVEGKGKAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 405
QY 254 ---QGTSGRGLGGQ-----ACAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 306
Db 406 GMLSGSGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 465
QY 307 A-----AAAAAAGAGAGAGAGAGAG-----327
Db 466 AVLINGGEGGICALAGKSGFGGGLLACADYNAPESTSPHNLQODILSFINEPTEA 525
QY 328 -----GSQFTSGPG-GYGPQQTSGRGLGGGAGAAAAAAGAG-----GOG 372
Db 526 LTRPLTNGDSTPGTGGDGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 584
QY 373 GYGLGSGQTSRGGGLGGGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 432
Db 585 GAGGAGGAGTAGAGGAGGAGGAG-----IGSGGTGGVGGAGATT-TGGVGGAGGAG 634
QY 433 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 487
Db 635 AGLLIGAAGLGGCGGAGTAAGTGGAGGTCGAAGLFGANGGAGGAGGAGGAGGAGGAGGAG 694
QY 488 YGGLGSGQ-GTSGPGGYGPQQTSGRGLGGGAGAAAAAAGAGAG-----QGGYGG 544
Db 695 AGGLYHGHTGGGNGGSGTGTAGTGGAGGPG-GLYAGGSGGAGGAGGAGGAGGAGGAGGAG 753
QY 545 GTS-----GRGLGGQ-----GAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 591
Db 754 AGSITLNASGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 813

```













[illegible]





Query Match		21.1%	Score 747.5;	DB 1;	Length 1356;
Best Local Similarity		30.6%	Pred. No. 2.4e-27;		
Matches 257;		Conservative 53;	Mismatches 353;	Indels 177;	Gaps 32;
QY	9	QMCGRSMASRGGLGGGAGAGAAAAA--AAAGGAG-----QGGYGGIGLSQGTSG-RG	58		
Db	215	ERCRAGPAGPAGAGADGSGTGPAGPLGAGPPGPPGAPGPKGEIGGAGSNGSPGPGQ	274		
QY	59	GLGGGAGAA-----AAAAAAGGAG-----QGGYGGIGLSQGTSG	94		
Db	275	GRCEPGINGAVGVPVGNPNNGINGCAKGAAGLPGVAGAPGPPGPRGGPGPOGSTG	334		
QY	95	RGGLGGGAGAAAAAAGAGAGGAGG--YGLGSGQTSRGGGLGGGAGAAAAAAG	153		
Db	335	ARGLGGDPGPGSQKGDGKAGPECHGVCVGAAGPAGEEGKRGSTGBVGATGPAGLRGARG	394		
QY	154	GAGGGYGGI-----GSQGTSGPGY--GPQQ-----TSGRGG	185		
Db	395	GAGTRGLPLEGRGGPIGMFGARGATGPGIRGAPDAGRAGESGLTGARGLPCNSGQGG	454		
QY	186	LGQGGAGAAAAAAGGAG-----QGGYGGIGLSQGTSGRGGGLGGGAGAAAAAAG	241		
Db	455	PPGKEGPPGAAGLDGRTGPPGTGPRGQPCNIGFPGKPGGAGKGDGKPTCATCLRG	514		
QY	242	GAGGGYGGI-----GSQGTSGRGGGLGGGAGAAAAAAGGAGAGGGYGG	288		
Db	515	GFADGNNGAPGAGVVGNTGKGEQGPAGPGFQLPLGPAGPAGEAGKAGNMGMPDQG	574		
QY	289	L-----GSQGTSG--RGLGGGAGAAAAAAGGAG--GGAG--GGYGGIGLSQGTSGPGG	337		
Db	575	LPGPAGVKERNSGPAGSAGSQCALGARGPAGTDPDGKGEPSVGVGAAGHQGGPG	634		
QY	338	YGPQQTSG-----RGLGGGAG--GAAAAAAGGAGQ-----GGY	374		
Db	635	M-PGERGAGGTPGPKCKEKGEGHRLGNMGRDARGGPGPGPPSGANGKEKESGSF	693		
QY	375	GGLSQGTSGRGGGLGGGAGAAAAAAGAGGAGG-----YGLGSGQTSRGGGLG	427		
Db	694	GPAGPAGLRGPGSGERGEGGAGLPAGPPGSDGQSGPRCKEPAGKGDVGPAGPAGPS	753		
QY	428	GCGAGAAAAAAGGAGGAGGAGG--GSQGTSGRGGGLGGGAGAAAAAAGGAGGAGG	486		
Db	754	GSGPSGAGPAGPPGGRDAGPSGLTGPPGAAGR--VGGPGP-AGIAGPPGAGPAGKD	810		
QY	487	GYGGL-----GSQGTSGPGYGPQQTSRGGGLGGGAGAAAAAAGGAGGAGG	536		
Db	811	GPRGLRGDPGPGPQGEQGVVGPAG-----ISGDKPSGESGPPGAPGAPGQVGLGPS	864		
QY	537	GYGGL-----GSQGTSGRGGGLGGGAGAAAAAAGGAGGAGGAGGAGGAGG	591		
Db	865	GFVGLPGSRGDKGLPGGPGAVGEPRLGPAGASGPRGPAAGNIGMPGNTGTQGEAGREGNS	924		
QY	592	GCGAGAAAAAAGGAGGAGGAGG--GYGGLGSGQTSG-----RGGLG--GOGAGA	635		
Db	925	GNDGPPRPGAGFKDRGEPSPGALGSSGQPGPNGPAGSAGRPONRSGSGPTGNGGPV	984		
QY	636	AAAAAAGGAG-----QGGYGGIGLSQGTSG-----PGYGGQQTSGIRIRAPS	681		
Db	985	GAVGARGAPCAPGPRGCKGAGEKGRGMKGLRGHGLQGMPCPNCPGSGTGSAGITGPA	1044		

Search completed: December 18, 2002, 16:18:15  
Job time : 21.6063 secs

Result No.	Score	Query			DB	ID	Description
		Match	Length	DB			
1	2012	56.8	644	5	Q8WSW4	Q8WSW4 nephila cia	
2	1998	56.4	617	5	Q46172	Q46172 nephila cia	
3	1729.5	48.8	544	5	Q46171	Q46171 nephila cia	
4	1594.5	45.0	691	5	Q9BIU3	Q9BIU3 dolomedes t	
5	1568	44.3	648	5	Q9BIU7	Q9BIU7 argiope tri	
6	1539.5	43.5	854	5	Q9BIU4	Q9BIU4 dolomedes t	
7	1452.5	41.0	1953	5	Q9BIU7	Q9BIU7 nephila mad	
8	1450	40.9	988	5	Q17434	Q17434 nephila cia	
9	1386	39.1	2639	5	Q76786	Q76786 antherea p	
10	1348.5	38.1	2655	5	Q964F4	Q964F4 antherea y	
11	1273	35.9	1217	16	Q8VIY9	Q8VIY9 mycobacteri	
12	1268.5	35.8	1715	16	Q8VIZ0	Q8VIZ0 mycobacteri	
13	1260.5	35.6	1489	16	Q53559	Q53559 mycobacteri	
14	1242	35.1	1079	16	Q53557	Q53557 mycobacteri	
15	1198.5	33.8	1306	16	Q53775	Q53775 mycobacteri	
16	1198	33.8	1384	16	Q8VIZ1	Q8VIZ1 mycobacteri	

```
Db 152 AGRGLGQGGAG---AAAAAAGGAGC-----GGLGGGAGGCGAGAGAAAAA 194
Qy 241 GGAGGGYGGGLGSGQTSRGGGLGGGAGAGAAAAAAGGAGGGYGGGLGSGQTSRGGGL 300
Db 195 GGAGGGYGGGLGSGQ-AGRGSGGGAGC-----AAAAAAGGAGGCGYGGGLGSGQ-AGRGGL 248
Qy 301 GGQGAGAAAAAAGGAGGCGYGGGLGSGQTSRGGYGP-CQQTSGRGGGLGGGAGAGAAA 359
Db 249 GGQGAG---AAAAAAGGAGGCGYGGGLGSGQ-AGQGGYGGGLGSGQAGRGGLGGGAG--- 301
Qy 360 AAAAAAGGAGGCGYGGGLGSGQTSRGGGLGGGAGAGAAAAAAGGAGGCGYGGGLGSGQ 419
Db 302 AAAAAAGGAGC-----GGLGGGAGGCGAGAAAAAAGGAGGCGYGGGLGSGQ 347
Qy 420 TSGRGGGLGGGAGAAAAAAGGAGGCGYGGGLGSGQTSRGGGLGGGAGAGAAAAAAGGAG 479
Db 348 -AGRGCG-----AAAAAAGGAGGCGYGGGLGSGQ-AGRGGLGGGAGC-----AAAAA 392
Qy 480 AGGAGGCGYGGGLGSGQTSRGGYGP-CQQTSGRGGGLGGGAGAGAAAAAAGGAGGCGY 538
Db 393 AGGAGGCGYGGGLGSGQ-AGQGGYGGGLGSGQSGRGGGLGGGAGC-----AAAAAAGGAG 444
Qy 539 GGLGSGQTSRGGGLGGGAGAGAAAAAAGGAGGCGYGGGLGSGQTSRGGGLGGGAGAGAA 598
Db 445 -----GGLGGGAGGCGAGAAAAAAGGAGGCGYGGGLGSGQ-AGRGGGAGC--- 487
Qy 599 AAAAAAAGGAGGCGYGGGLGSGQTSRGGGLGGGAGAGAAAAAAGGAGGCGYGGGLGSGQ 658
Db 488 --AAAAAAGGAGGCGYGGGLGSGQ-GRGLGGGAGC-----AAAAAAGGAGGCGYGGV 538
Qy 659 GTSRGGYGP-CQQTSGIRAPSTS 683
Db 539 ASAA-----SAAASRLSSPQAS 555

RESULT 2
O46172 ID O46172 PRELIMINARY; PRT; 617 AA.
AC O46172;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE Dragline silk protein spidroin 1 (Fragment).
OS Nephila clavipes (Orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
OX NCBI_TaxID=6915;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94165058; PubMed=8120021;
RA Beckwith R., Arcidiacono S.;
RT "Sequence conservation in the C-terminal region of spider silk
PT proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and Araneus
RT bicentenarius (Araneidae).";
RL J. Biol. Chem. 269:6661-6663(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98318805;
RA Beckwith R., Arcidiacono S., Stote R.;
RT "Evolution of repetitive proteins: spider silks from Nephila clavipes
RT (Tetragnathidae) and Araneus bicentenarius (Araneidae).";
RL Insect Biochem. Mol. Biol. 28:121-130(1998).
DR EMBL; U37520; AAC04504.1; --
FT NON_TER 1
FT NON_TER 617
FT NON_TER 617
SQ SEQUENCE 617 AA; 49410 MW; 1BFOCE9269A832E2 CRC64;

Query Match 56.4%; Score 1998; DB 5; Length 617;
Best Local Similarity 66.6%; Pred. No. 5e-101;
Matches 462; Conservative 18; Mismatches 58; Indels 156; Gaps 30;

Qy 2 ASWGGGQMGRCMSAGRGGLGGGAGAGAAAAAAGGAGGCGYGGGLGSGQTSRGGGLG 61
```

```
Db 3 AAAGAGGCGYGGGLGSGAGRGGQAG-----AAAAAAGGAGGCGYGGGLGSGQ-AGRGGLG 57
Qy 62 GCGAGAAAAAAGGAGGAGGCGYGGGLGSGQTSRGGGLGGGAGAGAAAAAAGGAGGAGGG 121
Db 58 GCGAG---AAAAAAGGAGC-----GGLGGGAGGCGAGAGAAAAAAGGAGGAGGG 100
Qy 122 YGGLGSGQTSRGGGLGGGAGAGAAAAAAGGAGGCGYGGGLGSGQTSRGGYGP-CQQTSG 181
Db 101 YGGLGSGQ-AGRGSGGGAGC-----AAAAAAGGAGGCGYGGGLGSGQ-----A 142
Qy 182 GRGGGLGGGAGAGAAAAAAGGAGC-----GAGQGGYGGGLGSGQTSRGGGLGGGAGAG 231
Db 143 GRGGGLGGGAGAGAAAAAAGGAGGCGYGGGLGGGAGGCGYGGGLGSGQ-AGRGGLGGGAG 200
Qy 232 AAAAAAAGGAGGCGYGGGLGSGQTSRGGGLGGGAGAGAAAAAAGGAGGCGYGGGLGSG 291
Db 201 --AAAAAAGGAGC-----GGLGGGAGGCGAGAAAAAAGGAGGCGYGGGLGSG 244
Qy 292 QGTSRGGGLGGGAGAGAAAAAAGGAGGCGYGGGLGSGQTSRGGYGP-CQQTSGRGGGL 350
Db 245 QG-----AGRGGGAGAGAAAAV-----GAGQGGYGGGQ-----AGQGGYGGGLGSGGAGRGGL 292
Qy 351 GGQGAGAAAAAAGGAGGCGYGGGLGSGQTSRGGGLGGGAGAGAAAAAAGGAGGAGGG 410
Db 293 GGQGAG---AAAAAAGGAGC-----GGLGGGAGGCGAGAAAAAAGGAGGAGGG 335
Qy 411 YGGLGSGQTSRGGGLGGGAGAGAAAAAAGGAGGCGYGGGLGSGQTSRGGGLGGGAG 470
Db 336 YGGLGSGQ-AGRGGGQ-----AAAAAAGGAGGCGYGGGLGSGQ-AGRGGLGGGAG 384
Qy 471 AAAAAAAGGAGGCGYGGGLGSGQTSRGGYGP-CQQTSGRGGGLGGGAGAGAAAAAAG 529
Db 385 ---AAAAAAGGAGGCGYGGGLGSGQ-AGQGGYGGGLGSGGSGRGGGLGGGAG-----AAAAA 435
Qy 530 AGGAGGCGYGGGLGSGQTSRGGGLGGGAGAGAAAAAAGGAGGCGYGGGLGSGQTSRGG 589
Db 436 AGGAGC-----GGLGGGAGGCGAGAAAAAAGGVRGGYGGGLGSGQ-----AG 477
Qy 590 LGGCGAGAAAAAAGGAGGCGYGGGLGSGQTSRGGGLGGGAGAGAAAAAAGGAGGAGG 649
Db 478 RGGGAG-----AAAAAAGGAGGCGYGGGLGGGCV-GRGGGLGGGAG-----AAAAAGGAGG 526
Qy 650 GYGGGLGSGQTSRGGYGP-CQQTSGIRIRAPSTS 683
Db 527 GYGGGCGSASNA-----SAAASRLSSPQAS 552

RESULT 3
O46171 ID O46171 PRELIMINARY; PRT; 544 AA.
AC O46171;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE Spidroin 1 (Fragment).
OS Nephila clavipes (Orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
OX NCBI_TaxID=6915;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90384959; PubMed=2402494;
RA Xu M., Lewis R.V.;
RT "Structure of a protein superfiber: spider dragline silk.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).
RN [2]
RP SEQUENCE OF 449-544 FROM N.A.
RX MEDLINE=98148687; PubMed=9487707;
RA Arcidiacono S., Mello C., Kaplan D., Cheley S., Bayley H.;
RT "Purification and characterization of recombinant spider silk
RT expressed in Escherichia coli.";
RL Appl. Microbiol. Biotechnol. 49:31-38(1998).
```















Qy	531	GGAGGGYGGGSGQTSG---	RGGLGGQ--GAGAAA-----	AAAAAAGGAGQ	573
Db	876	GAGNGGGGAGTGAAGLGDNGVGGDGGAGGAAGNGNAGVGLTAKAGDGAAGNGN			935
Qy	574	GGYGGGLSGQ----	TSRGGGLGQGA-----	GAIAAAAAAAGGA-----	CQGYGGLG 618
Db	936	GGAGGAGGAGDNNFNGGCGGAGGGQGGGLGGASTT			INANGAGNGGCTGKGAGGAG 995
Qy	619	SQSTSRGGGLGGGAGAAA-----	AAAAAAGGAGGGYGGGLGASQTS	SGPG-----	GYGP 668
Db	996	TLVGSGSGTGGDGDAGAGGGGFGGAAGKAGGGNGVGGDGGEGASGLGLGSLGFDG			1055
Qy	569	QQQTSQ 674			
Db	1056	GQGGQG 1061			
RESULT 13					
O53559 PRELIMINARY; PRT: 1489 AA.					
AC	O53559;				
DT	01-JUN-1998 (TREMBLrel. 06, Created)				
DT	01-JUN-1998 (TREMBLrel. 06, Last sequence update)				
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)				
DE	PGRS-family protein.				
GN	RV3514 OR RV023.21.				
OS	Mycobacterium tuberculosis.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.				
ON	NCBI_TaxID=1773;				
OX	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=H37RV;				
RC	MEDLINE=98295987; PubMed=9634230;				
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,				
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,				
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,				
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,				
RA	Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,				
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,				
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,				
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;				
RT	"Deciphering the biology of Mycobacterium tuberculosis from the				
RT	complete genome sequence."				
RL	Nature 393:537-544(1998).				
RL	EMBL; AL022022; CAA17751.1; -.				
DR	TubercuList; RV3514; -.				
DR	InterPro; IPR000084; PE_region.				
DR	InterPro; IPR002173; pFRB.				
DR	PF00934; PE; 1.				
DR	ProDom; PD001223; PE_region; 1.				
DR	PROSITE; PS00583; PFRB_KINASES_1; UNKNOWN_2.				
KW	Complete proteome.				
SQ	SEQUENCE 1489 AA; 115754 MW; 6855CBA1C3BAF3A CRC64;				
Query Match 35.6%; Score 1260.5; DB 16; Length 1489;					
Best Local Similarity 38.5%; Pred No. 6e-61;					
Matches 344; Conservative 35; Mismatches 274; Indels 241; Gaps					
Qy	6	GGQOMGRGSMASGRGLGQ-----	GAGAAAAA-----	AA 35	
Db	214	GCGAGGVGGAGGGTGGCAGGCGAGGAGTGGGATGGTGGHGGVGGDGGWLAP			273
Qy	36	AAAGGA-GGGYGGGLSQG--	TSRGGGLGGQ--GAGAAAAA	AAAAAAGGAGGCGGGGLGSGQ-	90
Db	274	GAGGAGGGGAGGAGSGDGAALGGTGGTGGTGGAGAGGCGGALLGAGCGGGLGAGCGG			333
Qy	91	-----	GTSGRGLGG-----	QGAAAAA	AAAAAAGGAGGCGGGGLGSGQTS 131
Db	334	GTGAGGDDVGLGVGTGKGVGVGAGLGGAGGAAGQ	LFASGAGNAGVGGAGGGQ--	---	391
Qy	132	GRGLGQGA-----	GAIAAAAAAAGGA-----	QGGYGGGLGSLGSGQTS	SGPGGY 175

RT	"Deciphering the biology of <i>Mycobacterium tuberculosis</i> from the	
RT	complete genome sequence."	
RL	Nature 393:537-544 (1998).	
DR	EMBL:	AL020202; CAA17749.1; -.
DR	TubercuList:	Rv3512; -.
DR	InterPro:	IPR002202; HMG-CoA_red.
DR	PROSITE:	PS00318; HMG_COA_REDUCTASE_2; UNKNOWN_1.
KW	Complete proteome.	
FT	NON_TER	1
SQ	SEQUENCE	1079 AA; A79718DCDB74B97D CRC64;

Query Match	35.1%	Score 1242;	DB 16;	Length 1079;
Best Local Similarity	40.7%;	Pred. No. 4.6e-60;		
Matches 316;	Conservative 34;	Mismatches 306;	Indels 120;	Gaps 29;
Qy	5	TGGQOMGRGSMASRRGGLGGOGAGANAAAAAAGGA-CGGGYG-----LGSQG	53	
Db	58	TGGQONGN--CGNGTGGKGTGGDGALAGSSGGAGKGGNGGDAGKAGTGSAPGAG	115	
Qy	54	TSRRGGLGGQ-----AGAAAAAAGGA-CGGYGGGLGSGTSGRGLGQ-	101	
Db	116	TGGDGGKGGGIGAAGTTPVGTGASGTTGSGGAGGTGDDGAANGTAGAGGNG	175	
Qy	102	-----GAGAAAAAAGACAGGGYGG-----LGSQTSRRGLGGOGAGAAA	145	
Db	176	GKGDGGAGVTSSTAGNSGGSGGKGDAGAGGAGATPGANGTAGNGDGD-	233	
Qy	146	AAAAAAGAGGGYGGGLGSO---GTSPPGYPGQQTSGRGLGGOGAGAAAAAA	201	
Db	234	AVGISGATGADGGHGTGAAGNGGTGGAGCGSIDVGCGTGTGTGGNGNGATGGAGG	293	
Qy	202	AGGAG-QGGYGGGLGSO-----CTSGRGLGGQCG--AGAAAAAAGAGAG	244	
Db	294	AGSGNSGGNGIGGKGGNAGAGGAAGSNGTGVANGTGGDGGGAAGATAAGSNGGAG	353	
Qy	245	QGGYGGGLGSGTSGRGLGGQ-----AGAAAAAAGAGG-----QGGY	286	
Db	354	TGSAGNG--GTGGGSGGAGGDDIGVGGKGGNGADGEVVGAGGAGSGPNTSPGNN	411	
Qy	287	GGLSGQSTSGRGLGGOGAGAAAAAAGAGAGGGYGGGLGSGQTSGP---GGYGPQ	343	
Db	412	GGQGGGSGGAG-CAAGAGCAGGANGTACNGCGGAGGTGGAGAASSATNGSGGAGG	470	
Qy	344	TSRRGGLGGQAGAAAAAAGACAGCGYGGGLGSGQTSRGLGGQ-----A	393	
Db	471	TGPDGSGGAGGTGGAGGTGGAAGDGGGGGGGAGG-CAGGQGGAGGAGGTGGNGNITG	529	
Qy	394	GAATAAAAAAGG-AGCGYGGGLGSO--GTSRRGLGGOGAGAAAAAAGAGAGGQGY	450	
Db	530	GTAGTAGAAGNGGAAGKGGAGCGCGGTGGTGGCGGAGDGGAGGTGGDRTVGGGTVPAGS	589	
Qy	451	GGLSGQSTSGRGLGGOGAGAAAAAAGAGAGGGYGGGLG-----SQGTSRPGYGP	505	
Db	590	GCQC--GNAGGGAGCG--GCADGSGDGDAGTGGNGNGNRSNGTGGAGGNGG	645	
Qy	506	QQTSGRGLGGOGAGAAA-AAAAAAGACAGCGYG-GLGSGQTSRGLGGOGAGAAAA	563	
Db	646	GANGGAGGAGSGSGGTGGNGGAGDADAGNCGNGNGTGNNGNGNGGAGGAGT	705	
Qy	564	AAAAAGACAGCGYGG-----GSGQTSRGLGGOGAGAAAAAAGAGAGCGYGGGLG	618	
Db	706	GSGNGCGSGGNGNAGMCGNSGTGSDGCGAGNGGAAGTCGTGGDGLTGTGGTGGSG	765	
Qy	619	SQTSRGLGGOGAGAAAAAAGAGAGCGYGGGLGSGQTSRPGYGPQQTSG	674	
Db	766	--GTGDCGCGNGCADNTANNTAAGDGGNGDGGFCGAGACGGGTTACAGCTG	819	

RESULT 15  
O53775  
ID O53775 PRELIMINARY; PRT; 1306 AA.  
AC O53775;  
DT 01-JUN-1998 (TREMBLrel. 06, Created)

DT	01-JUN-1998 (TrEMBLrel. 06, Last sequence update)	
DE	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)	
DE	PGRS-family protein (PE_PGRS family protein).	
GN	RV0578C OR MT0607 OR MT039.16C.	
OS	Mycobacterium tuberculosis.	
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;	
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.	
OX	NCBI_TaxID=1773;	
RP	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=H37RV;	
RC	MEDLINE=9825987; PubMed=9634230;	
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,	
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,	
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,	
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,	
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,	
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,	
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,	
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;	
RT	"Deciphering the biology of Mycobacterium tuberculosis from the	
RT	complete genome sequence.";	
RL	Nature 393:537-544(1998).	
RP	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=CDC 1551 / OSHKOSH;	
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,	
RA	Peterson J.F., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,	
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,	
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,	
RA	Bisbal W.;	
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and	
RT	laboratory strains.";	
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AL021942; CAA17449.1; -.	
DR	ENBL; AE006958; AAK44830.1; -.	
DR	TIGR; MT0607; -.	
DR	TuberculList; RV0578c; -.	
DR	InterPro; IPR000209; Peptidase_s8.	
DR	InterPro; IPR000084; PE_region.	
DR	InterPro; IPR002173; PfKB.	
DR	Pfam; PF00934; PE; 1.	
DR	ProDom; PD001223; PE_region; 1.	
DR	PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN_2.	
DR	PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.	
KW	Complete proteome.	
FT	CONFLICT 363 363 G -> R (IN REF. 2).	
SQ	SEQUENCE 1306 AA; 105964 MW; 843A30955FFA5686 CRC64;	

Query Match	33.8%;	Score 1198.5;	DB 16;	Length 1306;
Best Local Similarity	38.4%;	Pred. No. 1.2e-57;		
Matches 323;	Conservative	37;	Mismatches 293;	Indels 189; Gaps 34;

  

Qy	11	CRGSMA	SRGGLGGOGAGAAAAA	AAAAAGG	----	AGQGYG	GLGSQGT	SGRRGLGGOGAG	66							
		:	:	:	:	:	:	:	:							
Db	399	GTTST	SGNGDGGK	GADAISS	GQTG	ANGRG	DGGQV	NGCAG	--GAGRGAGGGLGF	456						
		:	:	:	:	:	:	:	:							
Qy	67	AAAAA	AAAAAGG	-AGQGYG	-	LGSGT	SRGGLGGOGA	-----	GAIAAAAAA	114						
		:	:	:	:	:	:	:	:							
Db	457	SEAP	PRPGAGGT	CGAGNGGT	QTAGD	GTGTG	GAGGAGD	GGSGAG	SGTGFNASAPCAAGSP	516						
		:	:	:	:	:	:	:	:							
Qy	115	GG	-----	AGQGYG	YGL	-----	GSQGT	SGRGLGGQA	-----	GAIAA	146					
		:	:	:	:	:	:	:	:	:						
Db	517	GGNG	NGPPG	CAGGEG	CAGGLA	ASGQ	NGSQG	AGDGG	AGNGGT	PCNGHGA	576					
		:	:	:	:	:	:	:	:	:						
Qy	147	--	AAAAA	GAGAC	GGYCG	GLSQGT	TSG	-PGYG	-PGQ	-QTS	SGRGLGGOGAGAAAAA	199				
		:	:	:	:	:	:	:	:	:	:					
Db	577	NGV	GAGHGG	DPV	CGAGG	CGGSG	STPG	ANGAP	NTPT	TS	GGNGNG	NGGRADAT	FGQT	636		
		:	:	:	:	:	:	:	:	:	:	:				
Qy	200	AAAG	-	-----	AGQGYG	GLGSQ	-----	GT	SGRGLGGOGAGAAAA	234						
		:	:	:	:	:	:	:	:	:						
Db	637	CAS	GRGDD	GG	LV	NGAG	GAGG	NGSK	GL	PGL	RLGN	PNPLD	GGT	TGNG	AGCGSGCAWAGN	696

Db 577 NCGVCGAGGGHGGPCVCGAGGGGGSGSTPGANGAPGNTPTSGGNGGNGGKGADATGFGGT 633  
 Qy 200 AAGG-----AGGGYGGGSG-----GTSRGSGGGGAGAAAA 234  
 Db 637 GASGRRGGDGLVNGGAGGAGGNGSKGLPGLGRLNPLDGGTGGCGGAGSGGGAGAGN 696



```
QY 235 AAAAAAGG-AGQGGYGGGLGSGQ-----GTSRGLGGGQ--AGAAAAAAA 277
    ||| ||| ||| |||
Db 697 GGTGGAGGTGGVGTGGSGSDGVNYSAGADGHPGGTGGVGTGGKGGDGDGGAAPNV 756
    ||| ||| ||| |||
QY 278 AGGAGQGGYGGGLGSGQTSRGLGGQAGAAAAAAG-GAQQGGYGGGLGSGQ-GTSGP 335
    ||| ||| ||| |||
Db 757 AGSQGGAGGSDG--GTGGYGGNGRGIDGADGATAGARQDGGAGGKGGRGGTGP 814
    ||| ||| ||| |||
QY 336 GGYGP-----GQTSRGLGGGQ-----AGAAAAAAA 364
    ||| ||| ||| |||
Db 815 GGAGPACTTGSQAGGNGSGGTGGDPDGGNGANGSVFTNNGIGGNGGNGAGPSPGAG 874
    ||| ||| ||| |||
QY 365 AAGGA-----GQGGYGGGLGSGQTSRGLGGQAGAAAAAAAAGGA 407
    ||| ||| ||| |||
Db 875 GSGGAGTFCATGSSSIHVNNGNGNGNGDHALSCNGAAGNGGNGCNRGSGCAG 934
    ||| ||| ||| |||
QY 408 GQGGYGGGLGSGQTSRGLGGQAGAAAAAAGGAGGQGGYGG-----LGSQGTSG 460
    ||| ||| ||| |||
Db 935 GHGNGGNASRGMGGDGTGGAGNAGQIGNGGAGNGDGTGSDGNPGAITGSGGRGG 994
    ||| ||| ||| |||
QY 461 RGLGGGAGAAAAAAGGAGGQGGYGGGLGSGQT-----SGPGGYGPGQQT 509
    ||| ||| ||| |||
Db 995 DGVGGQGGSVAGDGDGGRGA--GGTGTGLRGTGATCATCTFDAGADGHGNGGTG 1052
    ||| ||| ||| |||
QY 510 GRGLGGGAGAAAAAAA-----GGAGQGGYGGGLGSGQTSRGLGGGGA-G 558
    ||| ||| ||| |||
Db 1053 GVGGTGGAGGGGNGCAGGKALSPTGNNGSQAGGDDGAGCAG--GTGTTGGDGRGAHG 1110
    ||| ||| ||| |||
QY 559 AAAAAAAGGAGQGGYGGGLGSGQTSRGLGGQAGAAAAAAGAGQGGYGGLG 618
    ||| ||| ||| |||
Db 1111 TLFSSLAGTGTGTGNGGTGTG--GTGGAGGAGTGTGTLGATGATGAAGRAGNGVGGSG 1168
    ||| ||| ||| |||
QY 619 SQGTS-GRGLGGQAGAAAAAAGGAGQGGYGGLG-----SQGTSRGGY-GPGQQT 672
    ||| ||| ||| |||
Db 1169 GLGSFPGGTGGMG-GAGCTSTVSAGDGGRGFGDGLDASSGNGCDGHHGDFRT 1227
    ||| ||| ||| |||
QY 673 SG 674
    |||
Db 1228 AG 1229
```

***This Page Blank (uspto)***

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:05:10 ; Search time 35.254 Seconds  
(without alignments)  
1995.696 Million cell updates/sec

Title: US-09-490-291-8

Perfect score: 2700

Sequence: 1 MASMTGGQQMGRIRIRGYG.....GLSGCDVLQALLGHHHHH 528

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_101002:\*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2700	100.0	528	22	AA82611 Spider recombinant
2	2319	85.9	646	18	AAW27178 Nephila clavipes s
3	2196	81.3	718	19	AAW53346 Nephila clavipes s
4	2196	81.3	718	21	AAV59070 N. clavipes spider
5	2182	80.8	718	12	AAK14308 N.clavipes draglin
6	1999	74.0	651	20	AAV40097 Spider silk protei
7	1999	74.0	651	23	AAU11781 Spider natural sil
8	1996	73.9	1255	23	AAU50037 N. clavipes spidro
9	1996	73.9	1880	23	AAW50039 N. clavipes spidro
10	1970	73.0	989	23	AAW50038 N. clavipes spidro

11	1944	72.0	818	23	AAU11797 Dragline protein 1
12	1944	72.0	1626	23	AAU11798 Dragline protein 1
13	1919	71.1	630	23	AAW50042 N. clavipes spidro
14	1919	71.1	676	23	AAW50047 N. clavipes spidro
15	1914	70.9	809	23	AAU11793 Dragline protein 1
16	1914	70.9	1617	23	AAU11794 Dragline protein 1
17	1883.5	69.8	606	16	AAW50055 Spider dragline va
18	1883.5	69.8	606	20	AAV40102 Polymer of an anal
19	1879.5	69.6	606	20	AAV40101 Polymer of an anal
20	1871.5	69.3	604	16	AAW50057 Spider dragline va
21	1858.5	68.8	606	16	AAW50053 Spider dragline va
22	1858.5	68.8	606	20	AAV40100 Polymer of an anal
23	1784.5	66.1	691	22	AAW50055 Spider recombinant
24	1711.5	63.4	681	22	AAW50055 Spider recombinant
25	1623.5	60.1	777	23	AAW50046 N. clavipes spidro
26	1588.5	58.8	773	23	AAW50045 N. clavipes spidro
27	1587.5	58.8	364	23	AAW50043 N. clavipes spidro
28	1463	54.2	655	23	ABW76672 Protein related to
29	1344	49.8	831	16	AAW50035 PMISS1 MISP spider
30	1234	45.7	615	20	AAV40039 Spider silk protei
31	1209.5	44.8	271	23	AAW50044 N. clavipes spidro
32	1144	42.4	595	12	AAW14309 N.clavipes draglin
33	1144	42.4	595	19	AAW53347 Nephila clavipes s
34	1144	42.4	595	21	AAV59071 N. clavipes spider
35	1042.5	38.6	230	23	AAW50035 N. clavipes spidro
36	1008	37.3	291	22	AAW50035 Spider recombinant
37	973.5	36.1	465	22	AAW70188 Peptide dendrimer
38	972	36.0	1177	9	AAW50040 SLP III protein com
39	972	36.0	1177	11	AAW50307 SLP III (Silk-fibr
40	972	36.0	1177	17	AAW5105 Silk like protein
41	972	36.0	1177	18	AAW26342 Silk-like protein
42	972	36.0	1177	21	AAW78277 SLP III amino acid
43	972	36.0	1177	22	AAW72725 Repetitive protein
44	972	36.0	1177	22	AAW63995 SLP III protein seq
45	972	36.0	1178	14	AAW41007 Silk-like protein

#### ALIGNMENTS

RESULT 1

AAW50035

ID AAW50035 standard; Protein; 528 AA.

AC AAW50035;

XX 02-OCT-2001 (first entry)

DT Spider recombinant silk protein pETNCDs.

DE Spider; orb-weaver; silk protein; pETNCDs; structural protein;

KW purification; fibre; spinning.

XX Nephila clavipes.

OS Nephila clavipes.

PH Key Location/Qualifiers

FT Misc-difference 417 /note= "encoded by TAT"

FT Misc-difference 427 /note= "encoded by CGA"

FT Misc-difference 522 /note= "encoded by GAG"

XX WO200153333-A1.

XX 26-JUL-2001.

XX 01-NOV-2000; 2000WO-US30086.

XX 20-JAN-2000; 2000US-0490291.

XX (WELL)/ MELLO C M.

XX (ARCI)/ ARCIDIAONO S.



Db 60 GGCGAGAAAAAGGAGGAGGGLGGGAGGAG - AAAAAAGGAGGGYGLGSGAGRGGG 118  
 QY 61 AGAAAAAG-----CAGGGYGLGSGAGRGGLGGGAG--AAAAAGVGGGGL 108  
 Db 119 AGAAAAAGGAGGGYGGGAGGGYGLGSGAGRGGLGGGAGAAAAAGGAGGGGL 178  
 QY 109 GGCGAGGAGAAAAAGGAGGGYGLGSGAGRGGGGAGAAAAAGGAGGGYGL 168  
 Db 179 GGCGAGGAGAAAAAGGAGGGYGLGSGAGRGGGGAGAAAAAGGAGGGYGL 238  
 QY 169 GSGAGRGGLGGGAGAAAAAGGAGGGYGLGSGAGGGYGLGSGAGRGGLGG 228  
 Db 239 GSGAGRGGLGGGAGAAAAAGGAGGGYGLGSGAGGGYGLGSGAGRGGLGG 298  
 QY 229 GAG--AAAAAGGAGGGLG-----GCGAGAAAAAGGAGGGYGLGSGAGRGGAGAA 282  
 Db 299 GAGAAAAAGGAGGGGAGGGGAGGAGAAAAAGGAGGGYGLGSGAGRGGAGAA 358  
 QY 283 AAAAGGAGGGYGGGAGGGYGLGSGAGRGGLGGGAGAAAAAGGAGG----- 336  
 Db 359 AAAAG-----CAGGGYGLGSGAGRGGLGGGAG--AAAAAGGAGGGYGL 406  
 QY 337 -----GGLGGGAGAAAAAGGAGGGYGLGGGAGGAGGAGAAAAA 376  
 Db 407 GGCGAGGGYGLGSGGSRGGLGGGAGAAAAAGGAGGGYGLGGGAGGAG--AAA 462  
 QY 377 AAAGVRRGGYGLGSGAGRGGGAGAAAAAGGAGGGYGLGGGAGGAGGAG 436  
 Db 463 AAAGGAGGGYGLGSGAGRGGGAGAAAAAGGAGGGYGLGGGAGGAG 522  
 QY 437 AAAAGGAGGGYGGVGSASAAASRLSSPOASSRVSSAVSNLVASPTNSAALSSTI 496  
 Db 523 AAAAGGAGGGYGGVGSASAAASRLSSPOASSRVSSAVSNLVASPTNSAALSSTI 582  
 QY 497 SNVVSQIGASNPGLSGCDVLIQALL 521  
 Db 583 SNVVSQIGASNPGLSGCDVLIQALL 607

## RESULT 3

AAW53346

ID AAW53346 standard; Protein: 718 AA.

XX AC AAW53346;

XX DT 06-JUL-1998 (first entry)

XX DE Nephila clavipes spider silk protein.

XX KW Spider; Nephila clavipes; silk protein; tandem repeat; fibre; dragline;

XX KW cocoon; tensile strength; elasticity.

XX OS Nephila clavipes.

XX PN US5728810-A.

XX PD 17-MAR-1998.

XX PF 19-APR-1995; 95US-0425069.

XX PR 15-APR-1991; 91US-0684819.

XX PR 20-APR-1990; 90US-0511792.

XX PR 04-OCT-1994; 94US-0317844.

XX PR 19-APR-1995; 95US-0425069.

XX PA (UYWY-) UNIV WYOMING.

XX PI Hinman MB, Lewis RV, Xu M;

XX XX WPI; 1998-270437/24.

XX DR N-PSDB; AAW23249.

XX XX Recombinant spider silk proteins - useful for making fibres

PT

XX

PS

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ

Sequence 718 AA;

Query Match

Best Local Similarity 81.3%; Score 2196; DB 19; Length 718;

Matches 479; Conservative 4; Mismatches 10; Indels 188; Gaps 15;

The present sequence represents a spider silk protein from the present invention. Spider silk proteins, and peptide fragments within the proteins, can be produced and purified independently and can then be mixed and made into fibres that have higher tensile strengths and elasticity than naturally occurring fibres. The fibres can be used in mixed composites. The invention allows the two naturally occurring Nephila clavipes silk proteins to be produced independently so that they can later be combined to form silk fibres of high tensile strength and elasticity.

QY 6 GGQMGRIIRIRYGGGLGGGAGGAGAAAAAGGAGGGYGLGSGAGRGGAGGAGAA 65  
 Db 22 GGCGAG-----GGYGLGGGAGGAGAGAAAAAGGAGGGYGLGSGAGRGGAGGAGAA 78

QY 66 AAAGGAGGGYGLGSGAGRGGLGGGAG--AAAAAGVGGGGLG--OGAGGAG-GA 119  
 Db 79 AAAGGAGGGYGLGSGAGRGGLGGGAGAAAAAGGAGGGYGLGSGAGRGGAGGAG 138

QY 120 AAAAGGAGGGYGLGSGAGRGGGGAGAAAAAG-----CAGGGYGLG 166  
 Db 139 AAAAGGAGGGYGLGSGAGRGGLGGGAGAAAAAGGAGGGYGLGSGAGRGGAGG 198

QY 167 GLGSGAGRGGLGGGAGGAG----- 185  
 Db 199 GLGSGAGRGGLGGGAGGAGAAAAAGGAGGGYGLGSGAGRGGAGGAGGAGGAG 258

QY 186 -----AAAAAGGAGGGYGLGSGAGGGYGLGSGAGRGGLGGGAGGAGGAG 232  
 Db 259 GSGAGRGGAGGAGAAAAAGGAGGGYGLGSGAGGGYGLGSGAGRGGLGGGAGGAG 317

QY 233 AAAAGGAGGGYGLG-----GGAGAAAAAGGAGGGYGLGSGAGGAGGAGGAGGAG 284  
 Db 318 AAAAGGAGGGYGLGSGAGGAGGAGAAAAAGGAGGGYGLGSGAGRGGLGGGAGGAG 377

QY 285 AAGGAGGGYGLG-----CAGGG-----GYGLGSGAGRGGLGGGAGGAG 324  
 Db 378 AAGGAGGGYGLGSGAGRGGGAGAAAAAGGAGGGYGLGSGAGRGGLGGGAGGAG 437

QY 325 AA-----AAAAAGGAGG----- 336  
 Db 438 AAAAGGAGGGYGLGSGAGRGGGAGAAAAAGGAGGGYGLGSGAGRGGAGGAGGAG 497

QY 337 -----GGLGGGAGAAAAAGGAGGGYGLGSGAGRGGAGGAGGAGGAGGAGGAG 365  
 Db 498 AAAGGAGGGYGLGSGAGGGYGLGSGAGRGGLGGGAGGAGAAAAAGGAGGGYGLG 557

QY 366 QGAGAAAAAGGAGGGYGLGSGAGRGGGAGGAGAAAAAGGAGGGYGLGSGAGGAG 425  
 Db 558 QGAG-----AAAAAGGAGGGYGLGSGAGRGGGAGGAGAAAAAGGAGGGYGLG 613

QY 426 GAGGLGGGAGAAAAAGGAGGGYGLGSGAGAAAAASRLSSPOASSRVSSAVSNLVAS 485  
 Db 614 GRGLGGGAGAAAAAGGAGGGYGLGSGAGAAAAASRLSSPOASSRVSSAVSNLVAS 673

QY 486 PTNSAALSSTISNVVSQIGAS 506  
 Db 674 PTNSAALSSTISNVVSQIGAS 694

RESULT 4

AAW59070

ID AAW59070

XX XX AAW59070 standard; Protein: 718 AA.

XX AC AAW59070;











```

Db 940 GGLGGGAGGAGAAAAAGGAGCGGCGGGLGSGAGRGGLGGGAGAAAAAGGAGCGGY 999
QY 388 GGLGSGAGRGGAGAAAAAGGAGCGGCGGGLGSGAGRGGLGGGAGAAAAAGGAGCGGY 445
Db 1000 GGLGSGAGRGGAGAAAAAGGAGCGGCGGGLGSGAGRGGLGGGAGAAAAAGGAGCG 1059
QY 446 GGYGGVGS-----GASASAAA 462
Db 1060 GGYGGGSGAGRGGAGAAAAAGGAGCGGCGGGLGSGAGRGGLGGGAGAAAAAGGAGCG 1084

RESULT 10
AAM50038
ID AAM50038 standard; Protein; 989 AA.
AC
XX
XX
XX 18-SEP-2002 (first entry)
XX
XX N. clavipes spidroin synthetic homologue SOISM12 protein.
DE
XX Spidroin; spider; silk; fibre; film; membrane; wound; filter; SOISM12.
XX
XX Synthetic.
XX
XX DE10113781-A1.
XX
XX 13-DEC-2001.
XX
XX 21-MAR-2001; 2001DE-1013781.
XX
XX 09-JUN-2000; 2000DE-1028212.
XX
XX 24-OCT-2000; 2000DE-1053478.
XX
XX (IPKP-) IPK INST PFLANZENGENETIK & KULTURPFLANZE.
XX
XX Scheller J, Conrad U, Grosse F, Guehrs K;
PI
XX WPI; 2002-123561/17.
XX
XX N-PSDB; ABL61039.
XX
XX New DNA encoding synthetic spider silk protein, useful e.g. for closing
XX wounds, comprises modules that encode repeating units of spidroin
XX proteins -
XX
XX Claim 22; Page 34-38; 88pp; German.
XX
XX This invention describes a novel DNA sequence, encoding a synthetic
XX spider silk protein, comprising modules, each comprising a group of
XX sequentially arranged oligonucleotides, each oligonucleotide encoding
XX a repeating unit of a spidroin protein. The synthetic protein has at
XX least 84% homology with the Nephila clavipes spidroin protein and is
XX used to produce synthetic fibres, films and/or membranes, particularly: (i)
XX for medical use, especially to close wounds and/or to support or cover
XX artificial organs; (ii) as adhesion surfaces for culturing cells; and
XX (iii) as filters. The synthetic proteins are very similar to native
XX spider silk proteins; can be prepared on a large scale and can be spun to
XX fibres with excellent mechanical properties (strength and elasticity).
XX Also they retain water solubility after long-term boiling in aqueous
XX solutions and since they are also soluble in organic solvents but
XX precipitated at high salt concentration, they are easily extracted and
XX purified. The modular construction of the invention facilitates
XX incorporation of additional peptide-encoding sequences, e.g. to simplify
XX purification or modulate solubility. This sequence represents the
XX synthetic N. clavipes spidroin-1 homologue SOISM12 described in the
XX invention.
XX
XX Sequence 989 AA;
XX
XX Query Match 73.0%; Score 1970; DB 23; Length 989;
XX Best Local Similarity 68.6%; Pred. No. 5.5e-128;
XX Matches 428; Conservative 3; Mismatches 23; Indels 170; Gaps 15;

```

```

QY 6 GGOQMGRII-----RGYGLGGGAGCGGAGAAAAAGGAGCGGCGGGLGSGAGRGGQGA 61
Db 198 GGOQAGAAAAAGGAGCGGGLGGGAGCGGAG-AAAAAGAGAGCGGYGGLGSGAGRGQGA 256
QY 62 GAAAAAG-----GAGCGGCGGGLGSGAGRGGLGGGAG-AAAAAGCGVQGG 107
Db 257 GAAAAAGAGAGCGGCGGGLGGGAGCGGCGGGLGSGAGRGGLGGGAGAAAAAGGAGCGG 316
QY 108 LGGQAGAGAGAAAAAGGAGCGGCGGGLGSGAGRG----- 143
Db 317 LGGQAGAGAGAAAAAGGAGCGGCGGGLGSGAGRGGLGGGAGAAAAAGGAGCGGCG 376
QY 144 -----GSGGCGGAGAAAAAGGAGCGGCGGGLGSGAGRGGLGGGAG----- 185
Db 377 LGSQAGRGGCGGAGAAAAAGGAGCGGCGGGLGSGAGRGGLGGGAGAAAAAGGAGCGG 436
QY 186 ----- 198
Db 437 YGGLGSGAGRGGCGGAGAAAAAGGAGCGGCGGGLGSGAGRGGCGGAGAAAAAGGAGCGG 496
QY 199 YGGLGSGAGCGGCGGGLGSGAGRGGLGGGAG-AAAAAGGAGCGGGL-----CGGAGAAA 253
Db 497 YGGLGSGAGCGGCGGGLGSGAGRGGLGGGAGAAAAAGGAGCGGGLGGGAGCGGAGAAA 556
QY 254 AAAGGAGCGGCGGGLGSGAGRGGCGGAGAAAAAGGAGCGG----- 294
Db 557 AAAGGAGCGGCGGGLGSGAGRGGCGGAGAAAAAGGAGCGGCGGGLGSGAGRGGLGGGAG 616
QY 295 -----GGGCGGCGGCGGGLGSGAGRGGGLGGGAGCGGAGAAAAAGGAGCGG 337
Db 617 AAAAAAGGAGCGGCGGGLGGGAGCGGCGGGLGSGAGRGGLGGGAG----- 674
QY 338 GLG-----GCGAGAAAAAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 388
Db 675 GLGCGGAGCGGAGAAAAAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 734
QY 389 GLGSGAGRGGCGGAGAAAAAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 446
Db 735 GLGSGAGRGGCGGAGAAAAAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 794
QY 447 GYGGVGS-----GASASAAA 462
Db 795 GYGGGSGAGRGGCGGAGAAAAAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 818

RESULT 11
AAU11797
ID AAU11797 standard; Protein; 818 AA.
AC
XX
XX AAU11797;
XX
XX 26-MAR-2002 (first entry)
XX
XX Dragline protein 1 analogue DP-1B/his tag 8mer.
XX
XX Silk-like protein; SUP; transgenic plant; promoter; 5' terminator;
XX fabric production; material construction; rope; surgical suture;
XX flexible tie down; electrical component; implantation;
XX Dragline protein 1; DP-1B 8mer; His tag.
XX
XX Nephila clavipes.
XX Synthetic.
XX
XX WO200190389-A2.
XX
XX 29-NOV-2001.
XX
XX 24-MAY-2001; 2001WO-US16937.
XX
XX 25-MAY-2000; 2000US-206968P.
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
PA

```





Db 456 AAAA #59

||||

RESULT 14

AAM50047

ID AAM50047 standard; Protein; 676 AA.

XX AAM50047;

XX 18-SEP-2002 (first entry)

XX

XX N. clavipes spidroin synthetic homologue S01 protein #2.

DE

XX Spidroin; spider; silk; fibre; film; membrane; wound; filter; S01.

KW

XX Synthetic.

OS

XX

XX

FH Key Location/Qualifiers

FT Peptide 1..28

FT /label= LeB4\_signal\_peptide

FT Protein 29..659

FT /note= "Synthetic spidroin homologue S01"

FT Region 660..672

FT /note= "c-myc-tag"

FT Domain 673..676

FT /note= "ER retention signal"

FT

PN DE10113781-A1.

XX

XX 13-DEC-2001.

XX

XX 21-MAR-2001; 2001DE-1013781.

XX

PR 09-JUN-2000; 2000DE-1028212.

PR 24-OCT-2000; 2000DE-1053478.

XX

XX {IPKP-} IPK INST PFLANZENGENETIK & KULTURPFLANZE.

PA

PI Scheiller J, Conrad U, Grosse F, Guehrs K;

PI WPI; 2002-123561/17.

DR

XX

XX New DNA encoding synthetic spider silk protein, useful e.g. for closing

PT wounds, comprises modules that encode repeating units of spidroin

PT proteins

XX

PS Example 1; Fig 10A; 88pp; German.

XX

XX This invention describes a novel DNA sequence, encoding a synthetic

CC spider silk protein, comprising modules, each comprising a group of

CC sequentially arranged oligonucleotides, each oligonucleotide encoding

CC a repeating unit of a spidroin protein. The synthetic protein has at

CC least 84% homology with the Nephila clavipes spidroin protein and is used

CC to produce synthetic fibres, films and/or membranes, particularly: (i)

CC for medical use, especially to close wounds and/or to support or cover

CC artificial organs; (ii) as adhesion surfaces for culturing cells; and

CC (iii) as filters. The synthetic proteins are very similar to native

CC spider silk proteins; can be prepared on a large scale and can be spun to

CC fibres with excellent mechanical properties (strength and elasticity).

CC Also they retain water solubility after long-term boiling in aqueous

CC solutions and since they are also soluble in organic solvents but

CC precipitated at high salt concentration, they are easily extracted and

CC purified. The modular construction of the invention facilitates

CC incorporation of additional peptide-encoding sequences, e.g. to simplify

CC purification or modulate solubility. This sequence represents a construct

CC composed of the LeB4 signal peptide, N. clavipes spidroin-1 synthetic

CC homologue S01, a c-myc-tag and an endoplasmic reticulum (ER)-retention

CC signal described in the invention.

XX

SQ Sequence 676 AA;



GenCore version 5.1.3  
Copyright (C) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:14:14 : Search time 12.1566 Seconds  
(without alignments)  
1277.936 Million cell updates/sec

Title: US-09-490-291-8  
Perfect score: 2700  
Sequence: 1 MASMTGGQMGRIIRYGG.....GLSGCDVLIQLLHHHHHH 528

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2273	84.2	747	3	US-09-034-177-3
2	2196	81.3	718	1	US-08-425-069-2
3	2196	81.3	718	2	US-08-317-844B-2
4	1999	74.0	651	4	US-08-556-978B-19
5	1999	74.0	651	4	US-09-247-806-1
6	1883.5	69.8	606	4	US-08-556-978B-23
7	1883.5	69.8	606	4	US-09-247-806-8
8	1873.5	69.4	606	4	US-09-247-806-6
9	1871.5	69.3	604	4	US-08-556-978B-63
10	1858.5	68.8	606	4	US-08-556-978B-21
11	1858.5	68.8	606	4	US-09-247-806-4
12	1343.5	49.8	832	1	US-08-209-747-2
13	1343.5	49.8	832	1	US-08-458-298-2
14	1144	42.4	595	1	US-08-425-069-4
15	1144	42.4	595	2	US-08-317-844B-4
16	972	36.0	1136	4	US-08-806-029-9
17	972	36.0	1177	1	US-07-609-716-31
18	972	36.0	1177	1	US-08-175-155-29
19	972	36.0	1177	1	US-08-477-509B-64
20	972	36.0	1177	2	US-08-707-237A-35
21	972	36.0	1177	3	US-08-482-085B-64
22	972	36.0	1177	4	US-08-475-411A-31
23	972	36.0	1177	4	US-08-478-029A-31
24	972	36.0	1177	4	US-09-444-791A-64
25	970	35.9	714	4	US-08-556-978B-61
26	970	35.9	714	4	US-09-247-806-10
27	962.5	35.6	1059	1	US-08-175-155-48

28	962.5	35.6	1059	2	US-08-707-237A-54	Sequence 54, Appl
29	962.5	35.6	1059	4	US-08-806-029-10	Sequence 10, Appl
30	962.5	35.6	1101	1	US-08-477-509B-83	Sequence 83, Appl
31	962.5	35.6	1101	3	US-08-482-085B-83	Sequence 83, Appl
32	962.5	35.6	1101	4	US-09-444-791A-83	Sequence 83, Appl
33	929	34.4	1332	1	US-07-609-716-41	Sequence 41, Appl
34	929	34.4	1332	4	US-08-475-411A-41	Sequence 41, Appl
35	929	34.4	1332	4	US-08-478-029A-41	Sequence 41, Appl
36	922.5	34.2	1038	1	US-07-609-716-36	Sequence 36, Appl
37	922.5	34.2	1038	4	US-08-475-411A-36	Sequence 36, Appl
38	922.5	34.2	1038	4	US-08-478-029A-36	Sequence 36, Appl
39	921.5	34.1	766	1	US-08-175-155-53	Sequence 53, Appl
40	921.5	34.1	766	1	US-08-477-509B-88	Sequence 88, Appl
41	921.5	34.1	766	2	US-08-707-237A-61	Sequence 61, Appl
42	921.5	34.1	766	3	US-08-482-085B-88	Sequence 88, Appl
43	921.5	34.1	766	4	US-09-444-791A-88	Sequence 88, Appl
44	921.5	34.1	979	1	US-08-477-509B-89	Sequence 89, Appl
45	921.5	34.1	979	3	US-08-482-085B-89	Sequence 89, Appl

## ALIGNMENTS

RESULT 1  
US-09-034-177-3  
; Sequence 3, Application US/09034177  
; Patent No. 6127146  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: HUMAN FIBROUS PROTEIN  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/034,177  
; FILING DATE: HEREWITH  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0486 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 747 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: Genbank  
; CLONE: GI 1174414  
US-09-034-177-3  
Query Match 84.2%; Score 2273; DB 3; Length 747;  
Best Local Similarity 71.0%; Pred. No. 2.7e-160;























Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	1999	74.0	651	10	US-09-861-597-1	Sequence 1, Appli	
2	1883.5	69.8	606	10	US-09-861-597-8	Sequence 8, Appli	
3	1873.5	68.4	606	10	US-09-861-597-6	Sequence 6, Appli	
4	1858.5	68.8	606	10	US-09-861-597-4	Sequence 4, Appli	
5	970	35.9	714	10	US-09-861-597-10	Sequence 10, Appli	
6	915.5	33.9	529	10	US-09-861-597-2	Sequence 2, Appli	
7	679	25.1	357	10	US-09-864-761-35807	Sequence 35807, A	
8	649	24.0	1894	12	US-10-052-586-97	Sequence 97, Appli	
9	557	20.6	4679	10	US-09-804-898-2	Sequence 2, Appli	
10	534	19.8	3907	9	US-10-029-217A-24	Sequence 24, Appli	
11	529	19.6	450	10	US-09-812-382-6	Sequence 6, Appli	
12	524	19.4	720	10	US-09-756-071B-20	Sequence 20, Appli	
13	523.5	19.4	283	10	US-09-864-761-36720	Sequence 36720, A	
14	511	18.9	2211	9	US-10-096-961-1	Sequence 1, Appli	
15	500.5	18.5	201	10	US-09-848-990-22	Sequence 22, Appli	
16	500.5	18.5	201	10	US-09-760-364-14	Sequence 14, Appli	
17	499.5	18.5	200	10	US-09-798-584-18	Sequence 18, Appli	
18	499.5	18.5	200	10	US-09-967-624-19	Sequence 19, Appli	
19	499.5	18.5	200	10	US-09-988-667-18	Sequence 18, Appli	





Db 237 OQPSGPGSAAAAACPGQQGPGGYPGQQGPGGYGPGQQGPGSGPSAAAAAAGPGGY 296  
Qy 193 GAGQ---GGYG-----GLGQGAGQ---GGYG---GLGSGAGR 222  
Db 297 GPGQQGPGGYGPGQQGPGSGPSAAAAAAGPGGYGPGQQGPGGYGPGQQGPGGYGPGQ 356  
Qy 223 GGLGQGAGAAAAAG-----CAGQGLGQ-----QCAGAAAAAGCAGCGGYG 265  
Db 357 QGSPGPGSAAAAAAGPGQQGPGGYGPGQQGPGGYGPGQQGPGSGPSAAAAAAGPGGYG 416  
Qy 266 GLGSGAGRGQGG-----AGAAAAAGGAGQGGYGQGGAGQGGYGLGSGAGRGGLG 318  
Db 417 P-QQQGPGGYGPGQQGPGSGPSAAAAAAGPGGYGPGQQGPGGYG-P-QQQGPGGYGPG 474  
Qy 319 GQG-AGAAAAAAGAGQGGGLGQAG-----AAAAAG----- 352  
Db 475 QQPSGPGSAAAAAAGPGQQGPGGYGPGQQGPGGYGPGQQGPGSGPSAAAAAAGPGGY 534  
Qy 353 CAGQGLGQAGQ---CAGAAAAAAGGVRQGGY-----GLGSGQAG--- 396  
Db 535 GPGQQGPGGYGPGQQGPGSGPSAAAAAAG---PGGYGPGQQGPGGYGPGQQGPGGYG 591  
Qy 397 ---RGQCGAGAAAAAGGAGGTGGL-----GGQGVGAGGLGQCGAGAAAAAGCAG 446  
Db 592 PQQQPSGPGSAAAAAAGPGQQGPGGYGPGQQGPGGYGPGQQGPGSGPSAAAAAAGPG 651  
Qy 447 GYG-----GYGSGASASAAA 462  
Db 652 GYGPGQQGPGGYGPGQQGPGSGPSAAAAAAGAAA 684

RESULT 6  
US-09-861-597-2  
; Sequence 2, Application US/09861597  
; Patent No. US20020064539A1  
; GENERAL INFORMATION:  
; APPLICANT: PHILLIPPE, Michel  
; APPLICANT: GARSON, Jean-Claude  
; APPLICANT: ARRAUDEAU, Jean-Pierre  
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT  
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN  
; TITLE OF INVENTION: ANALOG  
; FILE REFERENCE: 6388-0365-0  
; CURRENT APPLICATION NUMBER: US/09/861,597  
; CURRENT FILING DATE: 2001-05-22  
; PRIOR APPLICATION NUMBER: 09/247,806  
; PRIOR FILING DATE: 1999-02-11  
; PRIOR APPLICATION NUMBER: FR 98/01614  
; PRIOR FILING DATE: 1998-02-11  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 529  
; TYPE: PRT  
; ORGANISM: Nephila clavipes  
US-09-861-597-2

Query Match 33.9%; Score 915.5; DB 10; Length 529;  
Best Local Similarity 47.9%; Pred. No. 2.4e-48;  
Matches 264; Conservative 20; Mismatches 136; Indels 131; Gaps 29;  
Qy 17 GYG-----GLGQGGAGQ-----GAG-AAAAAAGGAGQGGYG-----GLGSGQAGR-- 56  
Db 3 GYGPGQQGPGGYGPGQQGPGSGPSAAAAAAGPGGYGPGQQGPGGYGPGQQGPGGYG 62  
Qy 57 ---GGQAGAAAAAAGCAGGGYGLG--SQGAGRGGLGQGG-----AGAAAAAGVG 104  
Db 63 PQQQPSGPGSAAAAAAGSGQGGPGGYGPGQQGPGGYGPGQQGPGSGPSAAAAAAGS 122  
Qy 105 QGGLGQGGAGCAGAAAAAGAGGAGGGY----GLGSGAGRGSGGGQ--AGAAAAAGG 159  
Db 123 SGGPGGYGPGQ-----QQPGGYGPGQQGPGGYGPGQQGPGSGPSAAAAAAGS 171

Qy 160 AQQ-----GGYG-----GLGSGAGRGGLGQGGAGAGAAAAAAGCAGCGGYGGLGQGGAGQGGY 212  
Db 172 PQQQGPGGYGPGQQGPGGYGPGQQGPGSGPSAAAAAAGPGGYGPGQQGPGGYGPGQQGPGGY 231  
Qy 213 GGLGSGAGRGGLGQGGAGAAAAAGCAGGGYGLGQGGAG-----AAAAAAGCAGGG 263  
Db 232 -----GPGQGLSGPGSAAAAAAGPGQQGPGGYGPGQQGPGSGPSAAAAAAGPGG 285  
Qy 264 YG---GLGSGQAG--RGQGGAG-AAAAAAGGAGCGGYGQGGAGCGGYGGLGSGQAGRGGL 317  
Db 286 YGPGQQGPGGYGPGQQGPGSGAGSAAAAAAGPGQQGGLGGLGYPGQGGPGG---YGFQGGP 342  
Qy 318 GQCGAGAAAAAAGCAGCGGLGQAG-----AAAAAG-----GAGQGGGLG 360  
Db 343 GYGPGSASAAAAAAGPGQQGPGGYGPGQQGPGSGPSAAAAAAGPGGYGPGQQGPG 402  
Qy 361 GQAGQ---GAGAAAAAAGGVRQGGYGLGSGQAG-----RGQGGAG-AAAAA 409  
Db 403 GYAPGQGGPGSGPSASAAAAAAG--PGYGP-QQQGPGGYAPQQGPGSGPSAAAAA 459  
Qy 410 GGAGGGGTG---GLGGQGV-----GAGGLG-----GQGAGAAAAAGCAGCGGYG 449  
Db 460 ASAGPGGYGPAQQGPGSGPAGSASAGPGGYGPAQQGPGAGYGPAGYGPAGSASAGSAGY- 518  
Qy 450 GYVGSASASAA 460  
Db 519 GPGSQASAAAS 529

RESULT 7  
US-09-864-761-35807  
; Sequence 35807, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aecomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670



; PRIOR APPLICATION NUMBER: 60/081070  
; PRIOR FILING DATE: 1998-04-08  
; PRIOR APPLICATION NUMBER: 60/081195  
; PRIOR FILING DATE: 1998-04-09  
; PRIOR APPLICATION NUMBER: 60/081838  
; PRIOR FILING DATE: 1998-04-15  
; PRIOR APPLICATION NUMBER: 60/082568  
; PRIOR FILING DATE: 1998-04-21  
; PRIOR APPLICATION NUMBER: 60/082569  
; PRIOR FILING DATE: 1998-04-21  
; PRIOR APPLICATION NUMBER: 60/082704  
; PRIOR FILING DATE: 1998-04-22  
; PRIOR APPLICATION NUMBER: 60/082797  
; PRIOR FILING DATE: 1998-04-22  
; PRIOR APPLICATION NUMBER: 60/083322  
; PRIOR FILING DATE: 1998-04-28  
; PRIOR APPLICATION NUMBER: 60/083495  
; PRIOR FILING DATE: 1998-04-29  
; PRIOR APPLICATION NUMBER: 60/083496  
; PRIOR FILING DATE: 1998-04-29  
; PRIOR APPLICATION NUMBER: 60/083499  
; PRIOR FILING DATE: 1998-04-29  
; PRIOR APPLICATION NUMBER: 60/083559  
; PRIOR FILING DATE: 1998-04-29  
; PRIOR APPLICATION NUMBER: 60/084366  
; PRIOR FILING DATE: 1998-05-05  
; PRIOR APPLICATION NUMBER: 60/084414  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: 60/084639  
; PRIOR FILING DATE: 1998-05-07  
; PRIOR APPLICATION NUMBER: 60/084640  
; PRIOR FILING DATE: 1998-05-07  
; PRIOR APPLICATION NUMBER: 60/084643  
; PRIOR FILING DATE: 1998-05-07  
; PRIOR APPLICATION NUMBER: 60/085573  
; PRIOR FILING DATE: 1998-05-15  
; PRIOR APPLICATION NUMBER: 60/085579  
; PRIOR FILING DATE: 1998-05-15  
; PRIOR APPLICATION NUMBER: 60/085580  
; PRIOR FILING DATE: 1998-05-15  
; PRIOR APPLICATION NUMBER: 60/085582  
; PRIOR FILING DATE: 1998-05-15  
; PRIOR APPLICATION NUMBER: 60/085700  
; PRIOR FILING DATE: 1998-05-15  
; PRIOR APPLICATION NUMBER: 60/086023  
; PRIOR FILING DATE: 1998-05-18  
; PRIOR APPLICATION NUMBER: 60/086392  
; PRIOR FILING DATE: 1998-05-22  
; PRIOR APPLICATION NUMBER: 60/086486  
; PRIOR FILING DATE: 1998-05-22  
; PRIOR APPLICATION NUMBER: 60/087098  
; PRIOR FILING DATE: 1998-05-28  
; PRIOR APPLICATION NUMBER: 60/087208  
; PRIOR FILING DATE: 1998-05-28  
; PRIOR APPLICATION NUMBER: 60/087609  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/087759  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/087827  
; PRIOR FILING DATE: 1998-06-03  
; PRIOR APPLICATION NUMBER: 60/088025  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088028  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088029  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088033  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088167  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088202  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088212

; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088217  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088326  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088655  
; PRIOR FILING DATE: 1998-06-09  
; PRIOR APPLICATION NUMBER: 60/088722  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088738  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088740  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088811  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088824  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088825  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088826  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088861  
; PRIOR FILING DATE: 1998-06-11  
; PRIOR APPLICATION NUMBER: 60/088863  
; PRIOR FILING DATE: 1998-06-11  
; PRIOR APPLICATION NUMBER: 60/088876  
; PRIOR FILING DATE: 1998-06-11  
; PRIOR APPLICATION NUMBER: 60/089090  
; PRIOR FILING DATE: 1998-06-12  
; PRIOR APPLICATION NUMBER: 60/089105  
; PRIOR FILING DATE: 1998-06-12  
; PRIOR APPLICATION NUMBER: 60/089512  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089514  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089538  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089598  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089653  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089908

Query Match 24.0%; Score 649; DB 12; Length 1894;

Best Local Similarity 32.8%; Pred. No. 5.4e-32;

Matches 216; Conservative 12; Mismatches 230; Indels 200; Gaps 19;

Qy 26 ACQGAGAAAAAGGAGGGYGLGSGAGRGQGQAGAAAAAGGAGGGYGLGSGAG 85

Db 719 AGGGAAATTTAAAGGC-----TGAAAAAGAGAACCCAGAGGAGTAAGAA 768

Qy 86 R-----GGLGGGAGAAAA-----GGVGGGGLGSGAGQCA 117

Db 769 ATTGAAACCAAGGCACAAAAATTTTACTTTCTTTGGAGAGAG--CTGA 825

Qy 118 GAAAAAGGAGGGYGLGSGAGRGSGGGQAGAG-----AAAAAGGAGGGYGG 167

Db 826 GGAAGAAGAGGAGGAAGTAATCGAGTTAGTCAGACATGAAGGGCAAAAGCAAGTAG 885

Qy 168 LGSQG-----AGRGGLGGGAGAGAAAA 190

Db 886 TCATGACTTGTCTTAAGGATGCCACATCTCAGTTCTGTTCAGTTGTAGAAGTGA 945

Qy 191 AGGAG-----GGYGLGGQGA--GQGGYGLGSGAG----- 221

Db 945 AGGTGATGCCACGAGATTTAGTTGATGATGAGAGATGAAGTGCAGACATGATGA 1005

Qy 222 ----RGLGGQAGAAAAAGGAGGGYGLGSGAG-----AAAAAGGA 259

Db 1006 TATTGATGTTGATGAAGAACCTGATGAGAGAAAGATTGCCAAAAATTTAAAAAGCA 1065

Qy 260 -----GQGGYGLGSGAGRGQGQAGAAAAAGGAGGG 293

Db	501	CGCGACTTCTTGACGGAAATGCGCGGTGTGAGTAAGGCCCGGAGGCCCTTTCTTTTGTG	560
Qy	308	-----GSGAGRG-----	323
Db	561	CAATTTCAGAGGGAGAGACTACTTCCACATCAGCTGCTGCGTGAACACACCGGGGTG	620
Qy	324	AAAAAAAAGG-----AGGGGLGQ-----GACAAAAAAGG-----AGQG-----	357
Db	621	AAATCCATGGTTTGGGACGCTTTCCTGAGTCAGATTTCGCGAANAACGTATTCAGAGAAAT	680
Qy	358	----GLGCGGAGCGAG-----	386
Db	681	TACGCGCG-GATCAGCGCGACTTTGCCAAACTGGTTTCGCGGTCACAAGACCAGAAATGG	739
Qy	387	YGGLSGSGAGRGGGGAGAAAAAGGAGGGGTGLGGGGVGAGGLGG-----	432
Db	740	CGCCG---GAGGCGGCAACAA---GGTGGTGATGAGTGCTACATCCCCCAATTACTTCC	792
Qy	433	-----OQAGAAAAVGAGQG--GYGGVGSASASAASAAASRLSSPQASSRVS	475
Db	793	TCCCCAAAACCCAGCCTGAGCTCCAGTGGCGGTGGACTATATGGAAAC-----AGTATT	846
Qy	476	SAVSNLIVASGPTNSAALSSSTI-----SNVVSQIGASNPGLSGC	513
Db	847	TAAAGCGCTGTTTGAATCTCAGCGAGCGCTAAACGGTTGTGGTGC	899

RESULT 10  
US-10-029-217A-24  
; Sequence 24, Application US/10029217A  
; Patent No. US20020164735A1  
; GENERAL INFORMATION:  
; APPLICANT: OLSON, ERIC N.  
; APPLICANT: WANG, DA-ZHI  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO A CARDIAC-SPECIFIC  
; TITLE OF INVENTION: NUCLEAR REGULATORY FACTOR  
; FILE REFERENCE: UTSD:695US  
; CURRENT APPLICATION NUMBER: US/10/029.217A

```

: CORREMI FILING DATE: 2002-03-19
: PRIOR APPLICATION NUMBER: 60/2757,761
: PRIOR FILING DATE: 2000-12-21
: NUMBER OF SEQ ID NOS: 32
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 24
: LENGTH: 3907
: TYPE: PRT
: ORGANISM: Mus musculus
: US-10-029-217A-24

```

```

; SOFTWARE: PatentIn Ver. 2.1.1
; SEQ ID NO 24
; LENGTH: 3907
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-029-217A-24

Query Match          19.8%; Score 534; DB 9; Length 3907;
Best Local Similarity 31.3%; Pred.No. 7e-25;
Matches 201; Conservative 4; Mismatches 212; Indels 226; Gaps 26

Qy 19 GGLGSGAGQGGAGAAAAAGAGAGGGYGGGLGSGQAGR-----GGQGAGAA 64
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1060 GAAGTGGCAGAGCTGAAGCAG-GAGCTGAAGTTGCGATCACTGCTGCTCGGGCACC 1118
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 65 AAAAGGAGQGGYGGGLGSGQAGRGLGSGQAG-----AAAAGGAGGGGLGGQG 112
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1119 AAACGTGACTG-----ATTGAGCGCCTTCGAGCCTATCAAGACCAAAATCAGCCCTGTGCGCA 1174
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 113 AGQG-----AGAAAAAGGAGQGGYGGGLGSGQAGR 142
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1175 GGAGCCCCCAAGGCCCTCGCCCACTCTATCCTGCACNAGCCTGGCGAGGTG-----GT 1230
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 143 GSGGQGGAGAAAAAGAGAGQGGYGGGLGSGQAGRGG-----LGGQG---AGAAAAAAGG 193
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1231 GGTAGCCTTCCAGCGGGCCGSGCTGAGC-ACGGGGCCAGCCCTGTGTGGCAGCAGGCCCTG 1289
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 194 -----AGQGGYGGGLGGQG---AGQGGYGG---LGSQAGRGLGSGQAGAAAAAGAG--- 240
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1290 CTCACAGCTGAGGTGTGTTGGCCACCGGTGGCCAGCAGTGGGTGGTGAAGTTTGGCAGCA 1349
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Qy 241 -QGG-----LGGGAGAGAAAAAG--GAGQGGYGGIG 268  
Db 1350 CGGGCTCCACGCCCCCGTGTCTCCACCCCTCGAGCGCTCACTGCTCAGCAGCGGCG 1409  
Qy 269 SGGAG-----GGGAGAGAAAAAGAGQGGYGGGGA----- 299  
Db 1410 ATGAAACTCCACCCCGGGGACACCTTTTGGTAGATGGTGACATCACTCTACGCGAGC 1469  
Qy 300 -----GQGGYGGGSGQ-----AGRGGLGG-- 319  
Db 1470 TCACCTCGAGCCCTCGCCACTGCAGATCCTCGTGAAGCAGGAGGCCCCCGGGCCGGGT 1529  
Qy 320 -----QGAGAGAAAAAGAGAGQ-----GLGGQAGAGAAAAAGCA-GGGGLGG 361  
Db 1530 CCGTGTCTGAGCCCTGGGGGGCGGCGAGCTAGAGGGCGGCACAAGGACCAAGATGC 1589  
Qy 362 QGAGGAGAGAAAAAGAGYRQGGYGLGSGGA-----GRGGQAGAGAAAAAGAGAGGG 416  
Db 1590 TGCAGGAGAGAGAACAGCAGATCGA-GGGCTGAGCGCATGCTCCGCGAGAAGCAGCAG 1648  
Qy 417 -TGGGLGGGQV-----AGGLGGQ-AGAAAAVAGAGQ 446  
Db 1649 CTGGTGGAGCGCTCAACCTGCAGCTGAGCAGGAGAGAGCGAG 1691  
RESULT 11  
US-09-812-382-6  
; Sequence 6, Application US/09812382  
; Patent No. US20010034050A1  
; GENERAL INFORMATION:  
; APPLICANT: Chilkoti, Ashutosh  
; TITLE OF INVENTION: Fusion Peptides Isolatable by Phase Transition  
; FILE REFERENCE: 4176-101  
; CURRENT APPLICATION NUMBER: US/09/812.382  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/190,659  
; PRIOR FILING DATE: 2000-03-20  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 6  
; LENGTH: 450  
; TYPE: PRT  
; ORGANISM: ELP[V5A2G3-90]  
US-09-812-382-6  
Query Match 19.6%; Score 529; DB 10; Length 450;  
Best Local Similarity 35.9%; Pred. No. 2.2e-25;  
Matches 165; Conservative 16; Mismatches 234; Indels 44; Gaps 16;  
Qy 17 YGGLGGQAG-QGAGAGAAAAAGAGGGYGGGGLGSGQAGRGQAGAGAAAAAGGAGGGG 75  
Db 15 GVPGVGPGVPGAGVPGAGVPGGGVPG--GGV---PGGGVPGVPGVPGVPGVPGV 68  
Qy 76 YG--GLSQAGRGLGQAGAGAAAAAGVGGGGLGGQAGCGAGAGAAAAAGGAGQGGY 133  
Db 69 VGVPGVPGAGVPGAGVPGAGVPGGG--VPGGGVPGGGVPGVGV--PGVPGVPGVPGVPGV 125  
Qy 134 --GLSQAGRGGSGGQAGAGAAAAAGAGQGGY--GLSQAGAGRGGLGGQAGAGAAAA 189  
Db 126 VPGAGVPGAGVPGGGVPGGGVPGGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGAG 185  
Qy 190 AAGGAGGG----YGLGQAGAGQGGYGLGSGQAGRGGLGGQGA---GAAAAAGGAGGG 243  
Db 186 VPGGVPGGGVPGGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGV 245  
Qy 244 LGG---QGAGAGAAAAAGAGQGGY--GLSQAGRGGQAGAGAAAAAGGAGQGGYGGQ 298  
Db 246 VPGGGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGV 305  
Qy 299 AGQGGYGLGSGQAGRGLGGQAGAGAAAAAGAGGAGGGGLG---QGAGAGAAAAAGGAG 355  
Db 306 VPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGV 365

Qy 356 QGGLGSGAGQAGAGAGAAAAAGVRRGGYGGGLGSGQAGRGQAGAGAAAAAGGAGGQ 415  
Db 356 VPGVPGVPGV--PGAGVPGAGVPGGGVPGGGVPGGGVPGVPGVPGVPGVPGV----- 414  
Qy 416 GTGGLGGQGVGAGGLGGQAGAGAAAAAGAGQGGYGGVGG 454  
Db 415 GVPGVGPGVPGVPGAGVPGAGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGV 450  
RESULT 12  
US-09-756-071B-20  
; Sequence 20, Application US/09756071B  
; Patent No. US20020052307A1  
; GENERAL INFORMATION:  
; APPLICANT: Tryggvason, Karl  
; Kallunki, Pekka  
; Pyke, Charles  
; TITLE OF INVENTION: Laminin Chains: Diagnostic Uses  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fay Sharpe Fagan Minnich & McKee  
; STREET: 1100 Superior Ave, Suite 700  
; CITY: Cleveland  
; STATE: Ohio  
; COUNTRY: USA  
; ZIP: 44114  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/756.071B  
; FILING DATE: 08-Jan-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 09/663,147  
; FILING DATE: 150-September 2000  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Minnich, Richard, J.  
; REGISTRATION NUMBER: 24,175  
; REFERENCE/DOCKET NUMBER: TRV 20014  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 216-861-5582  
; TELEFAX: 216-241-1666  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 720 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:  
US-09-756-071B-20  
Query Match 19.4%; Score 524; DB 10; Length 720;  
Best Local Similarity 30.4%; Pred. No. 6.6e-25;  
Matches 210; Conservative 4; Mismatches 204; Indels 272; Gaps 29;  
Qy 19 GGLGQAGQAGAGAGAAAAAGAGQGGYGGGLGSGQAGRGQAGAGAAAAAGGAGGGYGG 78  
Db 12 GGAGGTTGCCAGTCAATA-----GGTTACTTTATGAGTTGCTAACCTGCTGAGCAGG 65  
Qy 79 -LGSQAGRG--LGGQAGAGAAAAAGVGGGGLGGQAGAGAGAAAAAGGAGGGYGG 134  
Db 66 AAGTTATGAGCAGGAG 121  
Qy 135 -----LGSQAGRGSGGQAGAGAAAAAGGAGGG----- 164  
Db 122 ACCCTAACTGGAGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 181  
Qy 165 -----YGLGSGQAGR--GGLGG-----QGAG-----A 185



Db 182 GATGTCCTTTAAAGCCGAGCTGATTCGGGCTGCTGCTGCTTATTTCTGAGTTAGGCTCTTTA 241  
QY 186 AAAAAAGG-----AGGGYGLGGGAG----- 208  
Db 242 AGATTGGCCCTCCCACTTTGAGGAGGGGGGCTGCTCTACCTCTGTGAATCTGCCCC 301  
QY 209 -----QGGYGLGSGAGRGG---LGQGA-----GAAAAAG-- 237  
Db 302 TGGACACCCCGGAGAGAGAGGCTCCGGGAATCTCGACATTCAGGCAAGGCT 361  
QY 238 ---GAGGGLGG-----QGAGAAAA-----AGAGGG----- 262  
Db 362 CCGGGCCGAGCCTCTGTGCCACACCCCTTGGCCCGGCGAGGTGTGCGCTCTCGCT 421  
QY 263 GYGLGSGAGRGGQAGAAAAAGGAGG-----GAG-----GAG- 300  
Db 422 GCGAGGGGAGGCGGGGCTGCGGGAGCGGATTTCCACCCCGTTTGTGCTGTGTGT 481  
QY 301 ---QGGYGLGSGAGR-----GGLGGGAGAAAAAGGAG--- 335  
Db 482 TTGCTGCTCTGGAGG-GCTGGGTCTCTTATTTCACAGGTGATGCACACCTGAAACA 540  
QY 336 -QGG-----LGGGAGAAAAAGGAGGGLGGGAGGGA-----GAAAAA 374  
Db 541 CAGGCTCTCTCTGTCAGGACTGAGTCAGGTAGAGAGTGCATATAAACCCACTGATCAA 600  
QY 375 AAAAGGVRGGYGLGSGAGRG-----CQAGAA---AAAAGAGGGTGGGCGGAG 428  
Db 601 GAAAAAGGAAGGACACAGCGGAGCGAGAGTGAACCAACCCAGGCGCGGCGAGCG 660  
QY 429 -----GLGGQG--AGAAAAVAGGQG 447  
Db 661 ACCCTGACGCGGAGACAGACTGAGCGG 690

RESULT 13

US-09-864-761-36720  
; Sequence 36720, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aemica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 36720  
; LENGTH: 283  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC005547.9  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 17  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 11  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.6  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9  
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 12  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 11  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 10  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.4  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12  
US-09-864-761-36720  
Query Match 19.4%; Score 523.5; DB 10; Length 283;  
Best Local Similarity 39.9%; Pred. No. 3.2e-25;  
Matches 139; Conservative 7; Mismatches 123; Indels 79; Gaps 13;  
QY 44 GYGYGLGSGAGRGGAGAAAAAGAGGGYGLGSGAGRGGLGGGAGAAAAAGGV 103  
Db 2 GGGGGGSDGGGGD-----GGGSDGGGGSDGGGGGGSDG-----GGG 42  
QY 104 GGGGLGGGAGGAGAAAAAGGAGGGYGLGSGAGRGGGGGGAGAAAAAGAGG 163  
Db 43 GGGGGGSDGGGGG--SDGGGGSDGGGGGGSDG---GGSGSDGG-----GSGDGG 91  
QY 164 GYGLGSGAGRGGLGGGAGAAAAAGAGGGYGLGSGAGGGYGLGGGAGGGYGLGSGAGRG 223  
Db 92 GGGGGGSDGGGGGGSDGG-----GGGSDGGSDGGSDGGGGGGSDGG-- 140  
QY 224 GLGGGAGAAAAAGAGGAGGGLGGGAGAAAAAGAG--OQGYGGLGSGAGRGCGAG 280  
Db 141 --GGDG-----DGGGGDGGSDGYGGRSDDGGGGGGSCSDGGGGGGSDGGGG 192  
QY 281 AAAAAAGGAGGGYGGGAGGGYGLGSGAGRGGLGGGAGGAAAAAGAGAAAAAGAGG 340  
Db 193 -----GGGSDGGGSDGGGGSDGDCGGGGSDRGG---CGSDGGGGSDGGG 238  
QY 341 GQAGAAAAAGAGGGGGLGGGAGGAGAAAAAGAGAAAAAGAGGVRGGY 388  
Db 239 GDGGG--SHGGGGGSDGGGGGGDAGGAD-----HDGGYG 273

RESULT 14

US-10-096-961-1  
; Sequence 1, Application US/10096961  
; Patent No. US20020155572A1  
; GENERAL INFORMATION:  
; APPLICANT: GUEGLER, Karl et al.  
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHOLIPASE PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHOLIPASE  
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF  
; FILE REFERENCE: CL000849DIV  
; CURRENT APPLICATION NUMBER: US/10/096,961  
; CURRENT FILING DATE: 2002-03-14











QY 492 LSSTISNVVSQIGAS--NPGLSG 512  
DB 859 FGGNTTSGASGGGNGGNACTAG 881

RESULT 9  
A70869  
hypothetical glycine-rich protein Rv2490c - Mycobacterium tuberculosis (strain H37Rv)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: A70869

A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: A70869

A; Status: preliminary; nucleic acid sequence not shown; translation not shown  
A; Molecule type: DNA

A;Residues: 1-1660 <COL>  
A;Cross-references: GB:AL021246; GB:AL123456; NID:g3261507; PIDN:CAA16067.1; PID:g2729  
A;Experimental source: strain H37Rv  
C;Genetics:

C; Superfamily: collagen alpha 1(IV) chain

Query Match 32.28; Score 868.5; DB 2: Length 1660;

Base local similarity	41.56	Prod. NO.	3.98	40					
Matches	245	Conservative	22	Mismatches	175	Indels	149	Gaps	30

Query Match	32.2%	Score 868.5;	DB 2;	Length 1660;
Best Local Similarity	41.5%	Pred. No. 3.9e-40;		
Matches 245;	Conservative 22;	Mismatches 175;	Indels 149;	Gaps

Qy 17 GYGGLGGGAGQGAGCAAAAAAAGAGGCGGYGGLGSQ ----GAGRCGGAGCAAAAAAGGAG 72  
 | | | | | : | | | | | : | | | | |  
 Db 649 GKGGAGGAGAGTEGYTGATGATVHSGGNGGKGGNGADATVAGANGCKGGAGSNGGLVGDGG 708

Qy	73	QGGYGGIGSGCA-----	-----GRGIGGCGCAGAAAAAGG----	102
		:	:	
Db	709	AGDGGGGAACANGNVYDGADGT	LGSGPGESEANGCGGCGVGGGACGAGDGCAGSS	768

**Qy**    103 -VGGGGCGGC-AGQGAGA-AAAAAGA-----CQQGVYGLGSQA--GRGGSG 146  
         : ||| !!: || | ||||  
**Db**    769 ALGGGNNGRGDAQAGGACGAGGAGGAGSVGDGPCKGKAGAGAGAGAGCGGGKG 828

**Qy**      147 GQGAGAAAAAGGACGGCGYCGLSQSGAGRGLGCQCAGAAAAGAAGGACGGYGCGLCGCG 206  
          || : | | | | : | : | | | : | | | | | | | | | | | | | | | | | | | | | | | |  
**Dd**      829 ASGADSAENVGGAGKGGDGGVGVG -GDGPPEGDG -GAGGAAPAGQVSHGYGVGGVDG 886

QY	207	-----ACQGCGGGLSGCA-----GRGLGG-----QCAGAAAAG-----GACQGG	243
Db	887	LGGAGNGGDGHGSDGGDGGDDGPAGLGGGLGDCSNGTTRAAAGSYDASDHGPFSSG	946

QY 244 LGGCGA-GAAAAAGGA-GGGTGG-----LGSOGA-GRGCQGAGAAAAAGGA----- 289

Db 947 NGGNGGCAQAASVAGGAGGNGGGNACRVDCGAGCNGCDGAAGANGANSAPGSDALA 100

**QY** 290 -GQ-GGYGGG-----AGQGGYGGLSGQAGRGCLGQ-CAGAAAAAAGGACQG 338  
|| || |||| | | | | : | | | | | | | | | |  
**Db** 1007 LGPPGGNGGQDAGACGAGGAGCGSVSDGGAGNGGAGGNGVGCSSGAGARG 106

QY 339 LGG---QGAGAAAAGGAGGGGLGGCGAGCAGCAAAAAA---GCVRGGVCGLG 391  
| | | | | | | | | | | : | | | |  
Db 1067 ANGIDSIGTGGAGGGGGDGGAGGVGGHGGDGVGGAAPSGTVGSHTGTGV---GGDCGLG 112

Db 1125 GAGGCGAGGCGGIGTGGAGGAGGNGDPCGAGRGGLGGDSNGTSAANGVDAASKHGP 1184  
Qy 426 ---GAGGLGGGAGAAAAGV-AGQGGYGG-----VSGSAAASAAA 462  
Db 1185 LTGGDGGVGGNCAKAAAAGDGGGCGDGNAGLFGDGGAGGADGTAEEA 1235  
RESULT 10  
S01820  
glycine-rich cell wall protein 1.8 precursor - kidney bean  
C:Species: Phaseolus vulgaris (kidney bean)  
C:Date: 30-Sep-1989 #sequence\_revision 19-May-1994 #text\_change 16-Jul-1999  
C:Accession: S01820  
EMBO J. 7, 3625-3633, 1988  
A:Title: Glycine-rich cell wall proteins in bean: gene structure and association of the  
A:Reference number: S01820; MUID:89091109; PMID:3208742  
A:Accession: S01820  
A:Molecule type: DNA  
A:Residues: 1-465 <REL>  
A:CROSS-references: EMBL:X13596; NID:g21002; PIDN:CAA31932.1; PID:g21003  
C:Comment: This protein is enriched in the cell wall fraction of young hypocotyls and ov  
tyls.  
C:Comment: Much of the sequence consists of tandemly repeated 22-residue segments with t  
C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8  
C:Keywords: cell wall; structural protein; tandem repeat  
F:1-30/Domaln: signal sequence #status predicted <SIG>  
F:31-465/Product: glycine-rich cell wall protein 1.8 #status predicted <MAT>  
Query Match 32.1%; Score 866.5; DB 1; Length 465;  
Best Local Similarity 47.8%; Pred. No. 1.8e-40;  
Matches 218; Conservative 8; Mismatches 177; Indels 53; Gaps 19;  
Qy 23 GCGAGCGAGAAAAAAGGAGCGGCGGLGSGQAGRGQAGAAAAAAGAGCGGCGGLGSGQ 82  
Db 33 GYGLHGTGGYGGAGSYGGCGGGGGGGGCGGYAGEHVYVGGSGGGGCGGCGGVTGGDQ 92  
Qy 83 GAGRGGLGQAGAAAAGGAGGCGGGLGQAGCGAGGAGAAAAAAGGAGCGGCGGLGSGQAGR 142  
Db 93 GAGYGGGGGGGGGAYVGGGGE--RGGYGGGCG-----GGAG-GGYGAGGEHIGY 141  
Qy 143 GSGGCGGAGAAAA--RAGNAGCGGCGGLGSGQAGRG-----LGGQAGAAAAAAGG 193  
Db 142 GGGGSGAGGGGYYNAGGAQGGGCGTGGAGGGGGGGGDDHGGGCGGCGGAGGCGGCGG 201  
Qy 194 AGCGGCGGLGCGGAGGCGGCGGLGSGQAGRGGLGQAGAAAAAAGAGCGGCGGLGSGAGAA 252  
Db 202 GGEHGGGGGGGGGAGGCGGCGGAGGCGGAG--GGCGGAGGCGGAGGCGG--GGAGCGGQ 258  
Qy 253 AAAGAGCGGCGGLGSGQAGRGCGGAGAAAAAAGGAGCGGCGGCGGCGGCGGLGSGQAG 312  
Db 259 GGAGGCGGAGGCGG--GGAGGCGGCG--GAGGCGGAGGCGGCGGCGGCGGCGGAGGCGG 316  
Qy 313 GRGGLGGCGAGAAAAAAGGAGCGGCGGLGQAGAGAAAAAAGGAGCGGCGGCGGCGGAG 370  
Db 317 GGGG--GGCGGAGGAGGAYAVGHHGGYGGGCGG-----DGGYGTGCGEHGGYGGGCGG- 370  
Qy 371 AAAAAAAGGVRQGGCGGLGSGQAGRGCGGAGAAAAAAG--CAGCGGTGGL-----420  
Db 371 --AGGCGTGGGHHGGY--GGCGG--GGYAGGCGHGAAGYGGGCGGCGGCGGCGGCGG 425  
Qy 421 --GGGCGAGGCGGAGAAAAAAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 454  
Db 426 HGGYGGGAGGGGCGGAGGAG--GGYGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 459  
RESULT 11  
A70896  
hypothetical glycine-rich protein Rv1091 - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 24-Nov-1999  
C:Accession: A70896

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Sqaers, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: A70896  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-853 <COL>  
A:CROSS-references: GB:AL021897; GB:AL123456; NID:g3256022; PIDN:CAA17207.1; PID:el25  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: Rv1091  
C:Superfamily: unassigned collagens  
Query Match 31.8%; Score 858; DB 2; Length 853;  
Best Local Similarity 42.8%; Pred. No. 8.4e-40;  
Matches 232; Conservative 18; Mismatches 184; Indels 108; Gaps 26;  
Qy 7 GQMGRIIRIRGYGGLGGGCGAGGAGAAAAAAGAGGAGCGGCGGLGSGQAG--CRGGGAGAG 63  
Db 292 GCGGSHALLWAGAGGCGGCGTGCAGGTAGAGNGGAGGGGTGGLFNGGAGGHHG 351  
Qy 64 AAA-----AAGGAGCGGCGGLGSGQAGRGGLGCG-----GAGAAAAAGCGGGL 108  
Db 352 AAAGNLAAGNGVSSGGGAGGTGGAGGCG--GAGGAGNARLWGVGGAGG--GDGGA 408  
Qy 109 GCGGAGCGAGAAAAAAGG-----GCGGCGGLGSGQ-----ACRGGSGGCGA----- 150  
Db 409 CGAGGCGGSLSGNANGCAGGDSRGCTGGAGGCGGAAGLLVGTGHHGGDGCAGGAATK 468  
Qy 151 --GAAAAAGGAGCGGCGGLGSGQAGRGGLGSGAGGAG-----AAAA 191  
Db 469 GGGGAAAGTGTAGAGGCGGAG-----GSGGSGDGGGGAAGPAGWLFPGDGCAGGNGGAAAA 524  
Qy 192 GGAG--QGGYGLGCGGCGGCGGCGGLGSGQAG-----GRGGLGCGGAGAAAAAAGG-- 239  
Db 525 GGAGGAGGCGGCGGNG--GNGGNGGNGGATGWLNGGAGGCGGATAGAGGANGVS 583  
Qy 240 -----GCGGLGCG--GAGAAAAAAGAGCGGCGGLG--SQAG--RGGCGAGAAAAAGG 288  
Db 584 STNGGCTGGNGGCTGGGCGGAGGAGLLGVGAGGCGGAGGCGGAGGCGGCGGCGGCGG 643  
Qy 289 AGCGGCGGCGGAGCGGCGGLGSGQAGRGGLGCG--GAGAAAAAAGGAGCGGCGGAGAG 347  
Db 644 A--GGDGGCG--CNGAGGTGGLFAGGNGGCGGCGGGAADICGNGAGGCG--GTGNGNG 699  
Qy 348 AAAGGAGCGGCGGCGGAGGAG-----GAAAAAAGGAGCGGCGGCGGCGGCGGCGGCGG 394  
Db 700 GSGGCGGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 759  
Qy 395 AGRGCGGA--GAAAAAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 450  
Db 760 GCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 819  
Qy 451 VG 452  
Db 820 AG 821  
RESULT 12  
B70812  
hypothetical glycine-rich protein Rv0834c - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: B70812  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Sqaers, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.



A;Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome  
A;Reference number: A70500; MUID:98295987; PMID:9634230  
A;Accession: B70812  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-882 <COL>  
A;Cross-references: GB:AL022004; GB:AL123456; NID:g3261550; PIDN:CAAL7640.1; PID:g291680  
A;Experimental source: strain H37Rv  
C;Genetics:  
A;Gene: RV0834c  
C;Superfamily: elastin

Query Match 31.8%; Score 857.5; DB 2; Length 882;  
Best Local Similarity 35.2%; Pred. No. 9.2e-40;  
Matches 247; Conservative 20; Mismatches 226; Indels 209; Gaps 29;

[illegible]

RESULT 13  
A70812

A70812  
 hypothetical glycine-rich protein RV0833 - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
 C:Accession: A70812  
 R;Cole, S.T.; Davies, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
 ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A:Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.

A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: A70812

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-749 <COL>

A:Cross-references: GB:AL022004; GB:AL123456; NID:g3261550; PIDN:CAAL7639.1; PID:g291

A:Experimental source: strain H37Rv

C:Genetics:  
A:Gene: Rv0833  
C:Superfamily: elastin

Query Match	31.5%	Score 850;	DB 2;	Length 749;
Best Local Similarity	36.5%	Pred. No. 2.1e-39;		
Matches 241;	Conservative 21;	Mismatches 233;	Indels 176;	Gaps 25;
QY	7	GOQGRIRIRYCGGLGCG-----CAGCGAGAAAAAAGGA-CQGGYGGGSGAGRG	57	
Db	20	GGPAGLLGVGGAGGAGDSAVAGVIGGAGGAGGNAALLFGAGGAGGAGSGSGGAAGGAGG	79	
QY	58	GQGAAAAAAGAGAGCGYGGGSGAG-----RGGLGGGAGAAA-----	98	
Db	80	AGGAGGLFASGSGCGFGPASTGTGGAGCTTCGACGLFASGVGVTGCGAGSGGTGCGVGT	139	
QY	99	-AAGV-----GQGLGGQAGCQAGAAAAAGGAGCGYGGGSGAGRGSGGQAGAA	153	
Db	140	GGAGGLFASGAGGAGSGGTGGAGTTCGAGGLFCAGGAGGLGCGNHTGCGHGCAGSAG	199	
QY	154	AAAGGAGCGGYGGLGSGQ-AGRGGLGQ-----CAGAAAAA-----AGG	193	
Db	200	LLALGDBGAGGAGGAATTGTGGAGGAGGKAGLLFGSGGAGGSGGAGCTFGDTGNSGGAGG	259	
QY	194	A-----GQGGYGGGLGGCGA-CQGGYGGGLGSGCARG--GLCGGQ--AGAAAAAGCAG	240	
Db	260	AGCKAGLLFGSGGAGSGGAGGFANGSTGGAGGAGGAGLIGNCGGSGGTSVATCGAG	319	
QY	241	QGGGSGGAGAA-----AAAGGAGGGYGGGL-----	267	
Db	320	NGGAGGAGGAGLITNGNGGSGCMGDPGCTGVGGTGGLLGLLDGANAPASTNPLHTAQ	379	
QY	268	-----GSQGA-----	288	
Db	380	QQAALAVNAPIQAVTRPLIGNGANGAPGSGAPGHHGWLFGGGTGCSSVSGAGGDDG	439	
QY	289	A-----GQGGYGGGAGCGYGGGLGSGCA-----GRGLGGGAGAAAAAAGGA--	334	
Db	440	AGGLLFCAGGAGGAGGAVTGTGATGSGSGAGGAGALLFCAGGAGGAGGSGGIGGFAAGCAG	499	
QY	335	GQGGGQGGAGAAAAAGGAG-CQGLGGGAGGAGCAG---AAAAAAGGAGVRGGYGG	390	
Db	500	GPGGAGGLFNGGAGGAGGAGSGVSGGACGEGGAGGAGGLFAGGAGGAGGSGNNVGGAGGA	559	
QY	391	GSQG-----AGRCQGAGAAAAAAGGAG-----	435	
Db	560	GGVGGGLFCAGGAGSGGGGVSAGDSGAGGNAGLLAPGLAGGAGGGGGGQGFDTGAGGPGG	619	
QY	436	GAAAAVGA-GGGYGGYGV-----SGASASAAASRL--SSPOASSRVSSAVSNLVSAGPTNSA	490	
Db	620	DAGLLVSGGAGGAGGFLTGTGPPCAAGGDAGLLFGSGGAGGAGGCGSRTDGLGAGGAGGK	679	
QY	491	A	491	
Db	680	A	680	

RESULT 14  
E70917

E/0517  
 hypothetical glycine-rich protein Rv1450c - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

[illegible]

RESULT 15  
F70963  
hypothetical glycine-rich protein Rv2634c - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 24-Nov-1999  
C:Accession: F70963

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
 ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998

A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno  
 A:Reference number: A70500; MUID:98295987; PMID:9634230  
 A:Accession: F70963  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-778 <COL>  
 A:Cross-references: GB:280225; GB:AL123456; NID:g3242265; PIDN:CAB02341.1; PID:c26639  
 A:Experimental source: strain H37Rv  
 C:Genetics:  
 A:Gene: Rv2634c  
 C:Superfamily: unassigned collagens

Query Match	31.3%;	Score 844;	DB 2;	Length 778;
Best Local Similarity	38.9%;	Pred. No. 4.5e-39;		
Matches	225;	Conservative	24;	Mismatches 204; Indels 126; Gaps 25;

  

QY	2	ASMTGGQ----	QMGRIRIRGCGGLGGOGAGQ--	AGAAAAAAGGA--	GQGSGYGGLSQ	53
		::	::	::	::	
Db	207	ATLVGTGTGGVGTATCLISGSFGGAGAAACVGTTGVGGSGVGVGFNGFSGAGSLG	266			
QY	54	AGRGOGAGAAAAAGAGQGYGGLSQGA-----	CRLGLG--	OQAGAAAAAGG----	102	
		::	::	::	::	
Db	267	AAGSVGAASYFTGTGGGVGGDGPAGDGAPGLITGNCGVGLGAGNAGNGGAGGM	326			
QY	103	-VGQGGILGGQ-----	AGCAGAAAAAGGA--	GQSGYGGLSQAG-----	141	
		::	::	::	::	
Db	327	LILGDGAGGQGPVAVAGVLGMPGAGNGGNANFNFGCGAGCGCTGLAGTNGYNPQSI	386			
QY	142	-----	RGSGGOGAGAAAAAGAGGQGYGGLSQAGRGLGGQ--	GA	183	
		::	::	::	::	
Db	387	NPNTGANCTDSNGNCNOTGGNGFPAGGVGEAGV--	CGQGLGESLDGNDGTGGKGA	444		
QY	184	GAAAAAAGGAGQGYGGLG-----	GQCAGQGYGGLSQ-----	AGRGLGGQCGAGA	232	
		::	::	::	::	
Db	445	GGTAGTDGAGGAGGAGGIGETDGSAGVATGTGGEGDGTGVDGVGGAGKGQGHNT	504			
QY	233	AAAAGGAGOCGLGQGCAGAAAAAGCAGQGYGGLSQ-----	-----	AGRCGGQAG	280	
		::	::	::	::	
Db	505	GVDAFDGDDGIGDGNALCAAGNGTGTCGAGNGRGMILINGAGGAGCTGGTGGG	564			
QY	281	AAAAAGCAGQGYGGQ-----	AGQGYGGLSQ-----	AGRGLGGQCGAGAA	326	
		::	::	::	::	
Db	565	GAAGFAGV--	GGAGBGLTDGATBEGTGLGLGVTGTGCMGSGGVGNGGAAGS	622		
QY	327	AAAAGCAGQGGLCGQGCAGAAAAAGCAGQGGLCGQGCAGCAGAAAAAAGAGVROGG	386			
		::	::	::	::	
Db	623	LILGGGGGAGVGGT-----	GTTGGTGGAGNGGAGGAGTTGGGATTGGGGGTGG	674		
QY	387	YGLLQSOG--	AGRCQ--	GAGAA-----	AAAAAGAGQGT--	433
		::	::	::	::	
Db	675	VGGAGTGTCTGAGGTTGSSGGAGLICWACAAGCTGAGTGTGGQGLCCQG--	GNGNGGT	733		
QY	434	GA-----	GAAAVAGQGGYGVGVSAGSAASAASRLSSP	468		
		::	::	::	::	
Db	734	GATGGCGDPAIG--	GNGAGGAGGSPGSGIOGNMGPP	771		

Search completed: December 18, 2002, 16:22:22  
Job time : 21.6009 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:06:04 ; Search time 8.10437 Seconds  
(without alignments)  
2702.183 Million cell updates/sec

Title: US-09-490-291-8

Perfect score: 2700

Sequence: 1 MASMTGGQGMGRIRINGGG.....GLSGCDVLIQALLGHHHHH 528

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2273	84.2	747	1 SPDI_NEPCL	P19837 nephila cla
2	1148	42.5	627	1 SPDI_NEPCL	P46804 nephila cla
3	952.5	35.3	5263	1 FBOH_BOMMO	P05790 bombyx mori
4	935	34.6	1901	1 Y208_MYCTU	O53553 mycobacteri
5	866.5	32.1	465	1 GRP2_PHAVU	P10496 phaseolus v
6	844	31.3	778	1 YQ34_MYCTU	P71933 mycobacteri
7	839.5	31.1	914	1 WA22_MYCTU	O06794 mycobacteri
8	828.5	30.7	641	1 EBN1_EBV	P03211 epstein-bar
9	819.5	30.4	801	1 Y747_MYCTU	O53810 mycobacteri
10	807	29.9	957	1 Y278_MYCTU	P56877 mycobacteri
11	782.5	29.0	603	1 YD25_MYCTU	O10637 mycobacteri
12	736	27.3	384	1 GRP1_PETHY	P09789 petunia hyb
13	729.5	27.0	860	1 ELS_MOUSE	P54320 mus musculu
14	714.5	26.5	338	1 LRP_ARATH	P27483 arabidopsis
15	697.5	25.8	481	1 LOR1_MOUSE	P18165 mus musculu
16	684	25.3	543	1 YP91_MYCTU	O50630 mycobacteri
17	680.5	25.2	864	1 ELS_RAT	O99372 rattus norv
18	671.5	24.9	498	1 Y118_MYCTU	O50615 mycobacteri
19	612	22.7	463	1 Y468_MYCTU	O50594 mycobacteri
20	609.5	22.6	515	1 Y140_MYCTU	O50594 mycobacteri
21	602	22.3	491	1 YK98_MYCTU	O10707 mycobacteri
22	595	22.0	747	1 ELS_BOVIN	P04985 bos taurus
23	593.5	22.0	750	1 ELS_CHICK	P07916 gallus gall
24	536.5	19.9	730	1 ELS_HUMAN	P15502 homo sapien
25	526.5	19.5	252	1 GRP1_PHAVU	P10495 phaseolus v
26	516	19.1	1733	1 VNUA_PRVKA	P33485 pseudorabie
27	512.5	19.0	672	1 PHX5_MOUSE	P08399 mus musculu
28	501.5	18.6	1356	1 CA21_ONCMY	O93484 oncorhynch
29	471.5	17.5	1355	1 CA21_RANCA	O42350 rana catesb
30	470	17.4	2038	1 FSH_DRONE	P13709 drosophila
31	455.5	16.9	966	1 FIB1_PETMA	P02674 petromyzon
32	454	16.8	316	1 LORI_HUMAN	P23490 homo sapien
33	439	16.3	671	1 CA11_RAT	P02454 rattus norv

ALIGNMENTS

RESULT 1

SPDI\_NEPCL

ID	SPDI_NEPCL	STANDARD;	PRT;	747 AA.
AC	P19837;			
DT	01-FEB-1991 (Rel. 17, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Spidroin 1 (Dragline silk fibroin 1) (Fragment).			
OS	Nephila clavipes (Orb spider).			
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;			
OC	Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.			
OX	NCBI_TaxID=6915;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE=90384959; PubMed=2402494;			
RA	Xu M., Lewis R.V.;			
RT	"Structure of a protein superfiber: spider dragline silk.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).			
RN	[2]			
RP	SEQUENCE OF 653-747 FROM N.A.			
RX	MEDLINE=94165058; PubMed=8120021;			
RA	Beckwith R., Arcidiacono S.;			
RT	"Sequence conservation in the C-terminal region of spider silk proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and Araneus bicentenarius (Araneidae).";			
RL	J. Biol. Chem. 269:6661-6663(1994).			
CC	-1- FUNCTION: Spiders major ampullate silk possesses unique characteristics of strength and elasticity. Fibroin consists of pseudocrystalline regions of antiparallel beta-sheet interspersed with elastic amorphous segments.			
CC	-1- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 2, OF THE DRAGLINE SILK.			
CC	-1- SUBCELLULAR LOCATION: Extracellular.			
CC	-1- DOMAIN: Highly repetitive protein characterized by regions of polyalanine and glycine-rich repeating units.			
CC	-1- SIMILARITY: BELONGS TO THE SILK FIBROIN FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M37137; AAA29380.1; -			
DR	EMBL; U03848; AAB60212.1; -			
DR	PIR; A36068; A36068.			
KW	Silk; Repeat.			
FT	NON_TER	1	1	25 X APPROXIMATE TANDEM REPEATS.
FT	DOMAIN	1	655	1.
FT	REPEAT	1	25	2.
FT	REPEAT	26	38	3.
FT	REPEAT	39	66	4.
FT	REPEAT	67	96	5.
FT	REPEAT	97	130	6.
FT	REPEAT	131	158	

34	434	16.1	1027	1	CAFF_RIPPA	P30754 riftia pach
35	432	16.0	1156	1	GLH4_CAEEL	O76743 caenorhabdl
36	429.5	15.9	1453	1	CA11_CHICK	P02457 gallus gall
37	428.5	15.9	1456	1	CA13_HUMAN	P02461 homo sapien
38	428	15.9	1464	1	CA11_HUMAN	P11087 mus musculu
39	427	15.8	1453	1	CA12_MOUSE	P28481 mus musculu
40	422	15.6	1459	1	CA11_MOUSE	O98917 canis famil
41	420	15.6	1460	1	CA11_CANFA	Q01119 mus musculu
42	419	15.5	1372	1	CA21_MOUSE	O46392 canis famil
43	415.5	15.4	1366	1	CA21_MOUSE	P02458 homo sapien
44	415	15.4	1418	1	CA12_HUMAN	P08123 homo sapien
45	412.5	15.3	1366	1	CA21_HUMAN	



```
Db 63 PQGGPSGSAASAAAGSCQGGPGYVPRQGGPGYVGGQGGPGSGPSAAASAAASAE 122
QY 104 -GQGLGGQAGGAGAGAGAGAGGGYV-----GLGSQAGRGSGGGGQ-AGAAAAA 157
Db 123 SGQGGPGYVGGQ-----GQGGYVGGQGGPGYVGGQGGPGSGPSAAAAA 171
QY 158 GGAGQ---GGYV---GLGSQAGRGGLGGGAGAGAGAGAGAGAGAGAGAGAGAG 210
Db 172 SGFGQGGPGYVGGQGGPGYVGGQGGPGSGPSAAASAAASGPGQGGPGYVGG 231
QY 211 GYGLGQAGRGGLGGQAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 261
Db 232 GY-----GPGQGLSGPSGSAASAAAGAGAGAGAGAGAGAGAGAGAGAGAG 285
QY 262 GYG---GLGSQAG--RGQGGAG--AAAAAGAGAGAGAGAGAGAGAGAGAGAG 315
Db 286 GYGPGQGGPGYVGGQGGPGSGAGSAAASAAAGAGAGAGAGAGAGAGAGAGAG 342
QY 316 GLGGQAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 358
Db 343 GPGYVGGPSASAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 402
QY 359 LGQGGAGQ-----GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 407
Db 403 PGYVPGQGGPGSGPSASAAASAAAG--PGYVPG--CQGGPGYVPGQGGPGSG 459
QY 408 AAGAGAGGGT---GLGGGV-----GAGGLG-----GGAGAGAGAGAGAG 447
Db 460 AAAAGPGYVPGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 519
QY 448 YGVGSGASASAAASRLSPSSAVSNLVASGPTNSAALSTISNVYSQICASN 507
Db 520 YG-----PGSASASASRLSPSGARVASVSNLVSSGPTSSAALSVISNV 575
QY 508 PGLSGCDVLQALL 521
Db 576 PGLSGCDVLQALL 589

RESULT 3
FB0H_BOMMO
ID FB0H_BOMMO STANDARD; PRT: 5263 AA.
AC P05790; Q26379; Q17220;
DT 01-NOV-1988 (Rel. 09, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fibroin heavy chain precursor (Fib-H) (H-fibroin).
GN FIBH.
OS Bombyx mori (silk moth).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC Diptera; Bombycoidea; Bombycidae; Bombyx.
OX NCBI_TaxID=7091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20330362; PubMed=10871375;
RA Zhou C.-Z., Confalonieri F., Medina N., Zivanovic Y., Esnault C.,
RA Yang T., Jacquet M., Janin J., Duquet M., Perasso R., Li Z.-G.;
RT "Fine organization of Bombyx mori fibroin heavy chain gene.";
RL Nucleic Acids Res. 28:2413-2419(2000).
RN [2]
RP SEQUENCE OF 1-168 FROM N.A.
RX MEDLINE=80045039; PubMed=498286;
RA Tsujimoto Y., Suzuki Y.;
RT "The DNA sequence of Bombyx mori fibroin gene including the 5'
RT flanking, mRNA coding, entire intervening and fibroin protein coding
RT regions.";
RL Cell 18:591-600(1979).
RN [3]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=79211211; PubMed=455439;
RA Tsujimoto Y., Suzuki Y.;
```

```
RT "Structural analysis of the fibroin gene at the 5' end and its
RL surrounding regions.";
Cell 16:425-436(1979).
RN [4]
RP PARTIAL SEQUENCE FROM N.A.
RX STRAIN=Kinshu X Showa;
RC MEDLINE=89094868; PubMed=3210244;
RA Mita K., Ichimura S., Zama M., James T.C.;
RT "Specific codon usage pattern and its implications on the secondary
RT structure of silk fibroin mRNA.";
RL J. Mol. Biol. 203:917-925(1988).
RN [5]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=94365842; PubMed=7916056;
RA Mita K., Ichimura S., James T.C.;
RT "Highly repetitive structure and its organization of the silk fibroin
RT gene.";
RL J. Mol. Evol. 38:583-592(1994).
RN [6]
RP SEQUENCE OF 5179-5263 FROM N.A., AND DISULFIDE BONDS.
RC STRAIN=J-139;
RX MEDLINE=99296390; PubMed=10366732;
RA Tanaka K., Kajiyama N., Ishikura K., Waga S., Kikuchi A., Ohtomo K.,
RA Takagi T., Mizuno S.;
RT "determination of the site of disulfide linkage between heavy and
RT light chains of silk fibroin produced by Bombyx mori.";
RL Blochm. Biophys. Acta 1432:92-103(1999).
CC -1- FUNCTION: FORMS THE SILK FILAMENT; A STRONG, INEXTENSIBLE,
CC INSOLUBLE AND CHEMICALLY INERT FIBRE.
CC -1- SUBUNIT: FORMED OF TWO CHAINS: HEAVY AND LIGHT, THAT ARE LINKED BY
CC A DISULFIDE BOND. HEAVY-LIGHT CHAIN ASSEMBLY IS ESSENTIAL FOR THE
CC EFFICIENT INTRACELLULAR TRANSPORT AND SECRETION OF FIBROIN.
CC -1- TISSUE SPECIFICITY: PRODUCED EXCLUSIVELY IN THE POSTERIOR (PSG)
CC SECTION OF SILK GLANDS.
CC -1- DOMAIN: COMPOSED OF ANTIPARALLEL BETA SHEETS. THE STRANDS OF THE
CC BETA SHEETS RUN PARALLEL TO THE FIBER AXIS. LONG STRETCHES OF SILK
CC FIBROIN ARE COMPOSED OF MICROCRYSTALLINE ARRAYS OF (-GLY-SER-GLY-
CC ALA-GLY-ALA-N) INTERRUPTED BY REGIONS CONTAINING HULKIER RESIDUES.
CC THE FIBER IS COMPOSED OF MICROCRYSTALLINE ARRAYS ALTERNATING WITH
CC AMORPHOUS REGIONS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AF226688; AAF76983.1; -
DR EMBL: V00094; CAA23432.1; -
DR EMBL: V00097; CAA23433.1; -
DR EMBL: S74439; AAB31861.1; -
DR EMBL: X13869; CAA32076.1; -
DR EMBL: M35378; AAZ27839.1; -
DR EMBL: AB017362; BAA33147.1; -
DR FIR: S01844; S01844.
KW Silk; Signal; Repeat.
FT SIGNAL 1 21
FT CHAIN 22 5263
FT DOMAIN 149 5206
FT DISULFID 5244 5244
FT DISULFID 5260 5263
FT CONFLICT 10 10
SQ SEQUENCE 5263 AA; 391586 MW; 8EE1D3A0A47440E CRC64;

Query Match 35.3%; Score 952.5; DB 1; Length 5263;
Best Local Similarity 43.3%; Pred. NO. 5.7e-37;
Matches 249; Conservative 28; Mismatches 201; Indels 97; Gaps 21;

QY 17 GYGGLGGQAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 65
DB 2691 GYGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2750
```



01-JUL-1989 (Rel. 11, Last sequence update)  
01-OCT-1994 (Rel. 30, Last annotation update)  
Glycine-rich cell wall structural protein 1.8 precursor (GRP 1.8).  
Phaseolus vulgaris (Kidney bean) (French bean).  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.  
NCBI\_TaxID=3885;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=cv. Tendergreen;  
MEDLINE=89091109; PubMed=3208742;  
Keller B., Sauer N., Lamb C.J.;  
"Glycine-rich cell wall proteins in bean: gene structure and  
association of the protein with the vascular system.";  
EMBO J. 7:3625-3633(1988).  
CC -1- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).  
CC -1- SUBCELLULAR LOCATION: Cell wall (Potential).  
CC -1- DOMAIN: THE REPEATED DOMAINS OF THE PROTEIN FORM A BETA-PLATED  
SHEET CONFIGURATION.  
CC -1- SIMILARITY: THE N-TERMINAL SIGNAL SEQUENCE OF ABOUT 30 AA SHOWS  
ABOUT 60% HOMOLOGY TO THAT OF THE GLYCINE-RICH CELL WALL  
STRUCTURAL PROTEIN GRP 1.0.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; X13596; CAA31932.1; --  
PIR; S01820;  
Cell wall; Structural protein; Repeat; Signal.  
FT SIGNAL 1 30 POTENTIAL.  
FT CHAIN 31 465 GLYCINE-RICH CELL WALL STRUCTURAL  
FT PROTEIN 1.8.  
FT DOMAIN 33 465 GLY-RICH.  
FT DOMAIN 205 359 8 X 22 AA TANDEM REPEATS.  
FT SEQUENCE 465 AA; 36683 MW; B5CA9B983B43607 CRC64;  
-----  
Query Match 32.1%; Score 866.5; DB 1; Length 465;  
Best Local Similarity 47.8%; Pred. No. 9.5e-34;  
Matches 218; Conservative 8; Mismatches 177; Indels 53; Gaps 19;  
QY 23 GQAGCGAGAAAAAAGAGGAGGGGGLGSGAGRGCGAGAAAAAAGAGGAGGGGLGSG 82  
DB 33 GYGLGHTGGYGAAGSYGGGGGGGGGGYAGGHHVYGGGGGGGGGGYGGDQ 92  
QY 83 CAGRGGLGCGAGAAAAAGGAGGGGGLGCGAGCGAGAAAAAAGAGGAGGGGGLGSGAGR 142  
DB 93 GAGYGGGGGGGGGGYAGGGG--RGYGGGG--GGAG-GGYGAGGHHGTGY 141  
QY 143 GSGGGGAGAAAAA-AAGAGGGGGLGSGAGRGG-----LGSGAGAAAAAAGG 193  
DB 142 GGGGGGAGGGGGYNAGCAGGGGTGTGGAGGGGGGGGGHGGGTCGGGAGGGGAGG 201  
QY 194 AGGGYGGGLGCGAGG--GGYGGGAGAGRGGLGCGAGAAAAAAGAGGAGGGGAGGAGAA 252  
DB 202 GGGGGGGGGGGAGGGYAGGAGGAGGAG--GGGAGGGAGGAGGAGGAGGAGGAGG 258  
QY 253 AAAAGGAGGGYGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 312  
DB 259 GGAGGGYAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 316  
QY 313 GRGGLGGGAGAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 370  
DB 317 GGGG-GGGGAGGGGGAAGVHHGGGGYGGGGGG-----DGGGGYGGGHHGGGGGGG 370  
QY 371 AAAAAAGAGGAGGGYGGGGLGSGAGRGCGAGGAGAAAAAAG--GAGGGGTGGL----- 420  
DB 371 --AGGGYGTGGEHGGG--GGGAGG--GGYAGGAGGAGGAGGAGGAGGAGGAGGAGG 425

QY 421 --GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 454  
DB 426 HGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 459

## RESULT 6

YQ34\_MYCTU STANDARD; PRT; 778 AA.

AC P71933;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical PE-PGRS family protein Rv2634c.  
GN RV2634c OR MT2712 OR MTCY441.04c.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteriia (class); Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
Baddock K., Basham D., Brown D., Chillingworth T., Connor R.,  
Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
Hornsbey T., Jagels K., Krogh A., McLean L.A., Moule S., Murphy L.,  
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
complete genome sequence.";  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDP 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
Delcher A., Utterback T., Weidman J., Khouli H., Gill J., Mikula A.,  
Bisshai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
laboratory strains.";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS  
SUBFAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; Z80225; CAB02341.1; --  
DR EMBL; AE007103; AAK47026.1; ALT\_INIT.  
DR TIGR; MT2712; --  
DR TuberculList; Rv2634c; --  
DR InterPro; IPR000084; PE\_region.  
DR Pfam; PF00934; PE; 1.  
DR ProDom; PD001223; PE\_region; 1.  
KW Hypothetical protein; Complete proteome.  
FT CONFLICT 51 51 V -> L (IN REF. 2).  
FT CONFLICT 63 63 Q -> H (IN REF. 2).  
FT CONFLICT 274 274 A -> T (IN REF. 2).  
SQ SEQUENCE 778 AA; 63131 MW; DAB20F5E8E4999E7 CRC64;  
-----  
Query Match 31.3%; Score 844; DB 1; Length 778;  
Best Local Similarity 38.9%; Pred. No. 1.4e-32;  
Matches 225; Conservative 24; Mismatches 204; Indels 126; Gaps 25;

QY 2 ASMTGQ-----QMGRIIRYCGGLGGGAGGQ--AGAAAAAAGGA-GGGYGGGLGSGQ 53  
 Db ATLVGTGGVGGATGLGSGGFGGAGGAAAGVCTTGCGVGGSGGVGGVFGNGFGGAGGLG 266  
 QY 54 AGRGGGAGAAAAAGGAGGGYGGGLGSGA-----GRGGLGG-OGAGAAAAAGG--- 102  
 Db AAGVGGAAAYFTGGGGVGGGCGAGCGDGCAGPLLLGNGGVGGLGAGAGANGGAGGM 326  
 QY 103 -VQGGGLGGGQ-----AGQGAGAAAAAGGA-GGGYGGGLGSGQAG----- 141  
 Db LLGDCGAGGGGPAVAGVLGMPAGGNGGNANWFGSGGAGGGGTGLAGTNGVPGSIA 386  
 QY 142 -----RGSQCGAGAAAAAAGGAGGGYGGGLGSGQAGRGGLGQ-GA 183  
 Db NPNTGANGTDSNGNGTGGNGGPGAGGVEAGGV--GGQGLGESLNDGTGKGGA 444  
 QY 184 GAAAAAAGGAGGGYGGGLG-----GQAGGGYGGGLGSGQ-----AGRGGLGSGGAGA 232  
 Db GGTAGTDCGAGGAGGAGGTGTTGDSAGGVATGEGDGTGAGVGGVGGAGGGGGGHN 504  
 QY 233 AAAAGGAGGGGLGGGAGAAAAAAGGAGGGYGGGLGSGQ-----AGRGGGGAG 280  
 Db GVGDAFGDGGIGGNGALGAAGNGGTGGAGNGGGRGMLGNGGAGGAGGTGCTGGG 564  
 QY 281 AAAAAGGAGGAGGGYGGQ-----ACGGYGGGLGSGQ-----ACRGGLGSGGAGAAA 326  
 Db GAAGFAGGV--GGAGGEGTUDGAGTREGTGGGLGGVGGTGGMGSGSGVGGNGGAAS 622  
 QY 327 AAAAGGAGGGGLGGGAGAAAAAAGGAGGGYGGGLGSGGAGAAAAAAGGAGVYRGG 386  
 Db LGLGGGGAGGGVGTG-----GIGGIGGAGNGGAGGAGTTGGGATGGGGTGG 674  
 QY 387 YGGLGSGQ--AGRGQ--GAGAA-----AAAGGAGGGT---GGLGGGAGGAGGLGGQ 433  
 Db VGGAGGTGTTGAGGTGGSGGAGGLGAGAGAGGTGAGTGGGAGGAGTGGGAGTGGG 733  
 QY 434 GA----GAAAVACAGGGYGGVCGSGASASASASASRLSSP 468  
 Db GATGGGGDFALG-GNGGAGGAGGSPGGSSGIGQNGMGP 771

RESULT 7  
 WA22\_MYCTU  
 ID WA22\_MYCTU STANDARD; PRT; 914 AA.  
 AC O06794;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE WAG22 antigen precursor  
 GN WAG22 OR RV1759C OR MT1807 OR MTCY28.25C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skellton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998)).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,

RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PCRS  
 CC SUBFAMILY.  
 CC -!- CAUTION: REF-2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A  
 CC FRAMESHIFT IN POSITION 85.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; Z95890; CAB09322.1; -;  
 DR EMBL; AE007040; -; NOT\_ANNOTATED\_CDS.  
 DR TIGR; MT1807; -;  
 DR TubercuList; RV1759c; -;  
 DR InterPro; IPR000084; PE\_region.  
 DR Pfam; PF00934; PE; 1.  
 KW Antigen; Repeat; Signal; Complete proteome.  
 FT SIGNAL 1 30 POTENTIAL.  
 FT CHAIN 31 914 WAG22 ANTIGEN.  
 SQ SEQUENCE 914 AA; 74354 MW; F6953C3DBE8E6AC8 CRC64;  
 Query Match 31.1%; Score 839.5; DB 1; Length 914;  
 Best Local Similarity 37.4%; Pred. No. 2.5e-32;  
 Matches 249; Conservative 22; Mismatches 210; Indels 185; Gaps 29;  
 QY 6 GGGQMGRIIRYCGGLGGGAGGAGAAAAA-----AGGAGQ-----GGYGGGLG 50  
 Db 129 GGDAGWLICNGGAGGSAKNGGAGGPGGAAGLFGNGGAGGAGGTATANNITGGAGGAG 188  
 QY 51 SQ-----GAGRGGGGAGAAAAAGG-----AGGGYGGGLGSGGAGRGGLG-----QGAGA 96  
 Db 189 GSANLFCAGGAGGAGGAATSLVGGTGTGTGNAGMLAGAGAGGAGGAGGAGGAGGAGG 248  
 QY 97 AAAAGGV-----GGGLGGGAGGAGGAGAAAAA-----CGAGGGYGGGLGSGGA--GRGSGGQ 148  
 Db 249 AGGAGGLFTTGGVGGAGGGHTGGGAGGAGGAGGLFCAGGAGGAGGFGDHTGLTGAGGAGD 308  
 QY 149 GAGAAAAAGGAGGGYGGGLGSGGAGRGGLGGGAGGAGAAAAAAGGAG-----QGG 198  
 Db 309 GGGGGGLFGAG--GDCGAGGSGLTGGAAGNGG--NAGTSLGAAAGGAGGTGGAGGTVF 365  
 QY 199 YGGLGSGGA-----GGGGYGGGLGSGGAGRGGLGGGAGGAGAAAAAAGGAGGGG 250  
 Db 366 KCGAGGAGGNAGMLFGSGGGGGTGGFGFAAGGGGGVGGGAGMLSGSGSGGAGGAGGAG 425  
 QY 251 AAAAAGGA-----GGGGYGGGLGSGGAGRGG--QGAGAAAAAAGGAGGGGGG--AG 300  
 Db 426 TAAGGAGGAGGAPGLTNGNGNGGSGGSGTGGVGGAGGAGNAVLTNGGEGIGALAGKSS 485  
 QY 301 QGGYGGGL-----GGGGYGGGLGSGGAGRGG--QGAGAAAAAAGGAGGGGGG--AG 314  
 Db 486 FGGFGLLGGADYNAPESTSPWHLQDDILSFINEPTTEALTRPLICNGDSGTFTGTD 545  
 QY 315 GGLGG-----QGAGAAAAAAGGA-----GGGLGGGAGGAGAAAAAAG 352  
 Db 546 GGAGGWLFGNGGNGAGAGTNGSAGGAGGAGGILFGTGGAGGAGGAGGTAGAGGAGGAG 605  
 QY 353 GA-----GGGGLGGGAGGQ-----CAGAAAAAAG-----GVRGGYGGGLG 391  
 Db 606 SAFLTGGSGTGGVGAATTTGGVGGAGGAGLIGAAGLGGCGGGAFTAGTTGGAGGTG 665  
 QY 392 -----SQAGRGGAGGAGAAAAAAGGAG-----QGGTGGGLGGGQ--VGAGGLGGQGA- 435



[illegible]



[illegible]

Db 219 GLFGAGGIGGAGGPFNG-GAGGAGGSRSLFEVLAAGGA--GGTGLSVNG-GTGGTGGT 274  
QY 182 GAGAA--AAAAAGGAG-----QGQGGYGLGSGQ-----AGRG 223  
Db 275 GGGGLFNSGAGGAGGFGVGSAGNGGTGGDGGTGTGNGGTGGTGTGNQLVGGEG 334  
QY 224 GLGGQAGAGAAAAGGAGGGLGQAGAGAAAAGGAGGQGGYGLGSGQA-----GRG 275  
Db 335 GAGGAGGNAIGLFGAGGIGGTGTGTGLG-APDPGGTGGKGVGGIGGAGALFPGGAGGTG 393  
QY 276 GOGAGAAAAAAGG-AGQGGYGGGQ--NGQGGYGLGSGQ-----GAGRGLGGQAGAA-AA 327  
Db 394 GFGASSADWAGIGGSGSGGNAKLIIDGAGGTGDSVGRGAGSGGTGGTGGTGLDGG 453  
QY 328 AAAAGGAG-----QGGLGGQAGAAAAAGGAGQGGGLGQAGAGAAAAAAGGV 382  
Db 454 AGGAGGTGTEFGSVGAGGAGG-GNAAGLSGAGGAGGAGGAGGAGGAGGAGGAGGAGG 512  
QY 383 ROGGYGLGSGQ-AGRGQAGAGAAAAGGAGGQGGTGGGLGQGVGA-GGLGGQAGAGAAA 440  
Db 513 GAGGAGGLGIADGGNGRG-----GKAGMVGNGSDGAGGASVAVANGVGGSGGNATLI 567  
QY 441 VGAGQGGYGGVG--GASASAAASRL-----SSPOAS 471  
Db 568 GNGNGGNGVGVSAPKGGAGGTAGLLGNGSPGLS 603  
RESULT 12  
ID GRP1\_PETHY STANDARD; PRT; 384 AA.  
AC P09789;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 01-AUG-1992 (Rel. 23, Last annotation update)  
DE Glycine-rich cell wall structural protein 1 precursor.  
GN GRP-1.  
OS Petunia hybrida (Petunia).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Solanales; Solanaceae; Petunia.  
OX NCBI\_TaxID=4102;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Condit C.M., Meadner R.B.;  
RT "A gene encoding a novel glycine-rich structural protein of petunia.";  
RL Nature 323:178-181(1986).  
CC -!- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).  
CC -!- SUBCELLULAR LOCATION: Cell wall (Potential).  
CC -!- MISCELLANEOUS: THIS PROTEIN CONTAINS 67% GLYCINE RESIDUES.  
CC -!- MISCELLANEOUS: 90% OF THE MATURE PROTEIN RESIDUES ARE CAPABLE OF  
FORMING A BETA-PLEATED SHEET COMPOSED OF 8 ANTI-PARALLEL STRANDS.  
CC -!- MISCELLANEOUS: THE GLYCINE-RICH REGION IS COMPRISED OF TWO RELATED  
FAMILIES OF REPEATS, F1 AND F2, EACH REPEAT CONTAINING ABOUT 40  
AA  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X04335; CAA27866.1; -.  
DR PIR; A26099; A26099.  
KW Cell wall; Structural protein; Repeat; Signal.  
FT SIGNAL 1 27  
FT CHAIN 28 384 GLYCINE-RICH CELL WALL STRUCTURAL  
FT PROTEIN 1.  
FT GLY-RICH.  
FT DOMAIN 41 384  
FT SEQUENCE 384 AA; 28777 MW; C8541C549417D18C CRC64;

Query Match 27.3%; Score 736; DB 1; Length 384;  
Best Local Similarity 46.8%; Pred. No. 7.5e-28;  
Matches 197; Conservative 10; Mismatches 130; Indels 84; Gaps 22;  
QY 40 GAGQGGYGLGSGQAGRG---GQAGAAAAAGGAGGQGGYGLGSGQAGRGGLGGQAGAGA 96  
Db 41 GCFGRRGGCGGGRFGRGPFSGRGRGAGGFGGGAGGGAGGGILG----GGGILGGGG--- 93  
QY 97 AAAAGGVGQGLGCGAGAGGAGAAAAAGGAGGQGGYGLGSGQAGRGSGGQAGAGAAAA 156  
Db 94 -GAGGGGLGGGCGAGGGFG-----GGAGGAGGGILG--CGGLGGGGCGGAGGGGV 143  
QY 157 AGGAGQGGYGLGSGQAGRGGLGGQAGAGAAAAAGGAGQGGYGLGSGQAGAGQGGYGLG 216  
Db 144 GGGAGSGGGFGAG-----GGVGG-CAGA-----GGGVGGGGFGG-----GGGGGVC 184  
QY 217 SQAGRGGLGGQAGAGAAAAAGGAGGQGGGLGGQAGAGAAAAAGGAGQGGYGLGSGQAGRG 276  
Db 185 G-GSGHGGGFGAGGGYGGGAGGLGGVG-----GGGGSGGGGIG-----GG 228  
QY 277 QGAGAAAAAGGAGQGGYGGQAGQGGYGLGSGQAGRGGLGGQAGAGAAAAAGGAGGAG 336  
Db 229 SCHGGCGFAGGGVG--CGVGGAGAGGGGGG--GGGGGGGGGLG--GSGHGGGFGAGGGVG 285  
QY 337 GGLGGQAGAAAAAGGAGQGGGLGGQAGQAGAGAAAAAGGAGYRQGGYGLGSGQAG 396  
Db 286 GAAGGVGGGGGFGGGGGGGVGGSGHGGFGAG-----GGVGGGAGGGILG----- 330  
QY 397 RGGQAGAAAAAGGAGQGGTGGGLGGQGVGAG--GLG--GQAGAGAAAVGAGQG--GYGGVGS 453  
Db 331 -GGGGAG-----GGGGIGGGHG--GGFGVGGIGVGAGAGAGGAGVGGVGGSGSGGGN 382  
QY 454 G 454  
Db 383 G 383  
RESULT 13  
EUS\_MOUSE  
ID ELS\_MOUSE STANDARD; PRT; 860 AA.  
AC P54320;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Elastin precursor (Tropoelastin).  
GN ELN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=BA1B/C; TISSUE=Lung;  
RX MEDLINE=95130069; PubMed=7829060;  
RA Wyder K.S., Sechler J.L., Boyd C.D., Passmore H.C.;  
RT "Use of an intron polymorphism to localize the tropoelastin gene to  
mouse chromosome 5 in a region of linkage conservation with human  
chromosome 7".  
RL Genomics 23:125-131(1994).  
CC -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND  
NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.  
CC -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER  
INTO AN EXTENSIBLE 3D NETWORK.  
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.  
CC -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:13:04 ; Search time 30.9992 Seconds  
(without alignments)  
3509.535 Million cell updates/sec

Title: US-09-490-291-8  
Perfect score: 2700  
Sequence: 1 MASMTGQQMGRIRIRGGY.....GLSGCDVLQALLGHHHHH 528

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriaph:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2555	94.6	544	5	O46171
2	2313	85.7	617	5	O46172
3	2290	84.8	644	5	Q8WSW4
4	1570.5	58.2	648	5	Q9BIU7
5	1405.5	52.1	691	5	Q9BIU3
6	1397	51.7	988	5	O17434
7	1305.5	48.4	854	5	Q9BIU4
8	1256	46.5	447	5	Q9BIV1
9	1099	40.7	636	5	Q16987
10	1042.5	38.6	563	5	Q9BIT5
11	1025.5	38.0	1953	5	Q9BIT7
12	1003	37.1	360	5	Q9BIU0
13	968.5	35.9	2639	5	Q76786
14	942.5	34.9	1715	16	Q8VIZ0
15	941	34.9	2655	5	Q964F4
16	935	34.6	1884	5	Q9NHW2

17	927	34.3	1079	16	O53557	O53557 mycobacteri
18	926	34.3	1217	16	O8VYI9	O8VYI9 mycobacteri
19	921	34.1	871	5	O44358	O44358 nephila cla
20	912.5	33.8	1489	16	O53559	O53559 mycobacteri
21	900.5	33.4	253	5	Q9BIT4	Q9BIT4 nephila sen
22	900	33.3	2249	5	Q9NHW4	Q9NHW4 nephila cla
23	893	33.1	1384	16	O8VIZ1	O8VIZ1 mycobacteri
24	892	33.0	1381	16	O53552	O53552 mycobacteri
25	890.5	33.0	738	5	O02402	O02402 pinctada fu
26	888.5	32.9	444	5	Q9BIU6	Q9BIU6 argiope tri
27	881	32.6	912	5	Q9BIT2	Q9BIT2 plectreureys
28	880	32.6	767	16	O53435	O53435 mycobacteri
29	877.5	32.5	1468	5	Q9GUB5	Q9GUB5 galleria me
30	873	32.3	1002	5	Q9BIU8	Q9BIU8 argiope tri
31	868.5	32.2	1685	16	O53215	O53215 mycobacteri
32	866.5	32.1	399	5	Q9BIT8	Q9BIT8 latrodectus
33	862	31.9	907	5	O44359	O44359 nephila cla
34	858	31.8	853	16	O53439	O53439 mycobacteri
35	858	31.8	1408	16	O8VK17	O8VK17 mycobacteri
36	857.5	31.8	882	16	O53845	O53845 mycobacteri
37	857	31.7	879	16	O8VKD2	O8VKD2 mycobacteri
38	852	31.6	651	5	Q9BIU9	Q9BIU9 argiope tri
39	850.5	31.5	233	5	Q9BIT6	Q9BIT6 nephila mad
40	850	31.5	749	16	O53844	O53844 mycobacteri
41	847	31.4	626	5	Q9NHW1	Q9NHW1 nephila mad
42	847	31.4	1329	16	O06810	O06810 mycobacteri
43	843.5	31.2	714	16	O53556	O53556 mycobacteri
44	836	31.0	1538	16	O53395	O53395 mycobacteri
45	835.5	30.9	1507	16	O8VJ23	O8VJ23 mycobacteri

ALIGNMENTS

RESULT 1

O46171	O46171	PRELIMINARY;	PRT;	544 AA.
AC	O46171			
DT	01-JUN-1998	(TrEMBLrel. 06, Created)		
DT	01-JUN-1998	(TrEMBLrel. 06, Last sequence update)		
DT	01-DEC-2001	(TrEMBLrel. 19, Last annotation update)		
DE	Spidroin 1 (Fragment).			
OS	Nephila clavipes (orb spider).			
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;			
OC	Araneomorphae; Entelegynae; Araneoldea; Tetragnathidae; Nephila.			
OX	NCBI_TaxID=6915;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=90384959; PubMed=2402494;			
RA	Xu M., Lewis R.V.;			
RT	"Structure of a protein superfiber: spider dragline silk.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).			
RN	[2]			
RP	SEQUENCE OF 449-544 FROM N.A.			
RX	MEDLINE=98148687; PubMed=9487707;			
RA	Aracidacono S., Mello C., Kaplan D., Cheley S., Bayley H.;			
RT	"Purification and characterization of recombinant spider silk			
RL	expressed in Escherichia coli.";			
RN	Appl. Microbiol. Biotechnol. 49:31-38(1998).			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98318805;			
RA	Beckwith R., Arcidiacono S., Stote R.;			
RT	"Evolution of repetitive proteins: spider silks from Nephila clavipes			
RL	(Tetragnathidae) and Araneus bicentenarius (Araneidae).";			
RN	Insect Biochem. Mol. Biol. 28:121-130(1998).			
RP	SEQUENCE FROM N.A.			
RA	Beckwith R.;			
RL	Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; U20329; AAC38957.1; -.			
DR	HSSP; P10969; IWGT.			
FT	NON_TER			









[illegible]

ID	Q16987	PRELIMINARY;	PRT;	636 AA.
AC	Q16987;			
DT	01-NOV-1996 (TREMBlrel. 01, Created)			
DT	01-NOV-1996 (TREMBlrel. 01, Last sequence update)			
DE	Major amputate spidroin 2 (Fragment).			
DE	Fibroin-3 (Fragment).			
GN	ADF-3.			
OS	Araneus diadematus (Spider).			
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;			
OC	Araneomorphae; Entelegynae; Araneidae; Araneus.			
OX	NCBI_TaxID=45920;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=96178678; PubMed=8600519;			
RA	Guerette P.A., Ginzinger D.G., Weber B.H., Gosline J.M.;			
RT	"Silk properties determined by gland-specific expression of a spider			
RT	fibroin gene family";			
RL	Science 272:112-115(1996).			
DR	EMBL; U47855; AAC47010.1; -.			
DR	InterPro; IPR001419; Glutinin.			
DR	PRINTS; PR00210; GLUTENIN.			
FT	NON_TER 1			
SQ	SEQUENCE 636 AA; 56074 MW; E2F41180C978BB59 CRC64;			
Query Match 40.7%; Score 1099; DB 5; Length 636;				
Best Local Similarity 52.7%; Pred. No. 8.4e-49;				
Matches 307; Conservative 20; Mismatches 158; Indels 98; Gaps 26;				
QY	7	GOQMRIRIRGYGLGGGCA-GGAGAAAAAAG--GAGGGYGLGSGAGRGQGAG 62		
DB	45	GOQSPQGGGPGGQGYGPGASAAAAAGGYGPGSGQGPGGQGY-----GPGSS 99		
QY	63	AAAAAGGAGGCGYGLGSGAGRGGLGGGAG---AAAAAGGVGQG-----GLGG 110		
DB	100	AAAAAGGNGPCS---GQGAGQGPGQGPGASAAAAAGGYGPGSQGPGQGPGG 155		
QY	111	QCA-GGAGAAAAAGGAGCGGGLGSGAGRGSGGGGCGAGAAAAAGGAGCGG-GYGL 168		
DB	156	QCPYGPASAAAAAGGYGPGSGQGPGQGPGGQGYGPGASAAAAAGGYGPGSQGQP 215		
QY	169	GSQAGRGGLGGGQ---AGAAAAAGGAGGQ-----GYGLGGQ-----AG 208		
DB	216	GOQPGQGGPGGQGYGPGASAAAAAGGYGPGYGOQPGQGPGGQGYGPGASAA 275		
QY	209	GGYGG-GLGSGAGRGGLGGQ-----AGAAAAAGGAGGGLGGGAGAAAAAGGAGQ 261		
DB	276	SGGYGPGSQGPGQGPGGQGYGPGASAAAAAGGYGPGS-GQGGPGQ-----GPGQ 329		
QY	262	GGYGLGSGAGRGCGGAGAAAAAG---GAGCGGYGCGGAGCGGY---GLGSGAGRG 315		
DB	330	QGPQGGPGGQGYGPGASAAAAAGGYGPGSGQGPGQGPGQGPGQGPGQG 389		
QY	316	GLGGGAGAAAAAAGGAGGAGGGLGGQ-----GAGAAAAAG---GAGGGLGGGAG- 365		
DB	390	GPGQGPGQ-----GPGQGPGQGYGPGASAAAGGYGPGSQGPGQGPGQG 442		
QY	366	QGAGAAAAAAGGVGQ-GYGLGSGAGRG---GCGAGAAAAAGGAGCGG-GTGGL 420		
DB	443	QCPG-----QCPGPGQGPGQGPGQGPGQGPGQGPGQGPGQGPGQGPGQG 494		
QY	421	CGQYVAGGLGGQ---AGAAAAAGCGGCGYGGVSGASNAASRLSPQASRRVSSAV 478		
DB	495	GOQPGQGPGGQGYGPGASAAVSVGGYGPQSSVPVASARLSRSPASRRVSSAV 554		
QY	479	SNLVSPTNSAALSTISNVYSQIGASNPGLSGCDVLIOALL 521		
DB	555	SSLVSSGPTKHAALNTISSVVSQVSNPGLSGCDVLIOALL 597		
RESULT 10				
Q9BIT5				
ID	Q9BIT5	PRELIMINARY;	PRT;	563 AA.
AC	Q9BIT5;			

DT	01-JUN-2001 (TREMBlrel. 17, Created)			
DT	01-JUN-2001 (TREMBlrel. 17, Last sequence update)			
DE	Major amputate spidroin 2 (Fragment).			
GN	MASP2.			
OS	Nephila madagascariensis.			
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;			
OC	Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.			
OX	NCBI_TaxID=115969;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21179804; PubMed=11283372;			
RA	Gatesy J., Hayashi C., Motriuk D., Woods J., Lewis R.;			
RT	"Extreme Diversification, Conservation, and Convergence of Spider Silk			
RT	Fibroin Sequences";			
RL	Science 291:2603-2605(2001).			
DR	EMBL; AF350278; AK30607.1; -.			
FT	NON_TER 1			
FT	NON_TER 563			
SQ	SEQUENCE 563 AA; 48361 MW; 0AEAD885304B8B6B CRC64;			
Query Match 38.6%; Score 1042.5; DB 5; Length 563;				
Best Local Similarity 49.9%; Pred. No. 5.6e-46;				
Matches 284; Conservative 21; Mismatches 155; Indels 109; Gaps 23;				
QY	29	GAGAAAAAAGGAGQ---GGYG-----GLGSGAGRGCGGCGAGAAAAAGGAGCGGY 76		
DB	5	GPGSAAAAAAGPGGCGGYGPGQGPGGPGSGSNAANAANAANAAGPGQGP 64		
QY	77	GGLGSGAGRGGLGGGAGAAAAAG-GYCGGLGGGCGGAGAAAAAGGAGCGGY--- 133		
DB	65	GG---YGPQPGGPGGYGPGQGPGSGYGPGQGPGSGPSAASA-AAAAAGCGCGCGY 120		
QY	134	--GLGSGAGRGSGGCGAGAAAAAGGA--GOGYVGLGSGGAGRGGLGGG---ACA 185		
DB	121	QCGPGYGPQGPGSGSAAAAAAGPGQGPGGYGPGQGPGYGP-GQGPGGYGPGQGPGPSA 179		
QY	186	AAAAAGGAGCGGYGGLGCGGAGCGGYGGLGSGAGRGGLGGGCGAGAAAAAGGAGCG 245		
DB	180	AAAAAAGPGGCGGYGPGQGPGGY-----GPGCGPSGPGSNAANAANAAGPGQGP 233		
QY	246	GCGAGAAAAAGCGCGGYGGLGSGAGRGCGGAGAAAAAGGAGCGGCGGAGCGGY 305		
DB	234	GYGPGQ-----QGPYGP-GQGPGS-GPGSAAAAAAGGAGCGCGYGPQ----- 280		
QY	306	GLGSGAGRGGLGGGAGAAAAAAGGAGCGGLG-----GOGAGAAAAAGG 353		
DB	281	----GPGCGGPGSGSAAAAAAGPGPGYGPQGPGYGPQGPGYGPQGPGSAAAAAAG 336		
QY	354	ACGGGLGGGAGQ-----GAGAAAAAAGGAGVROGGY----- 388		
DB	337	PGCGPGGYGPGQRPSPGYGPGQGPGSGPSAAAAAAGPGQPGAYGPGSGSNAANA 396		
QY	389	GLGSGG-AGCGCGAGAAAAAGGAGCGGTG-----GLGCG--VGAGGLG-----G 432		
DB	397	GLGGYGPAGGAGSAGSAAAAA-AAGPGGYGPGQGPGSGPSAAGPGYGPAGQGPARTY 455		
QY	433	QCGAGAAAVGAGCGGYGSGASASASRLSPQASRRVSSAVSNLVSPTNSAAL 492		
DB	456	PGSAAAAAAGSAGYG-----PGQASASASRLSPDSCARVASVSNLVSPTNSAAL 511		
QY	493	SSTISNVYSQIGASNPGLSGCDVLIOALL 521		
DB	512	SSVISNAVVSQIGASNPGLSGCDVLIOALL 540		
RESULT 11				
Q9BIT7				
ID	Q9BIT7	PRELIMINARY;	PRT;	1953 AA.
AC	Q9BIT7;			
DT	01-JUN-2001 (TREMBlrel. 17, Created)			
DT	01-JUN-2001 (TREMBlrel. 17, Last sequence update)			
DT	01-JUN-2001 (TREMBlrel. 17, Last annotation update)			



```

RN  SEQUENCE FROM N.A.
RP  MEDLINE=20496953; PubMed=11040284;
RA  Sezutsu H., Yukuhiro K.;
RT  "Dynamic rearrangement within the antheraea pernyi silk fibroin gene
RI  is associated with four types of repetitive units.";
RL  J. Mol. Evol. 51:329-338(2000).
DR  EMBL; AF083334; AAC32606.1; -.
SQ  SEQUENCE 2639 AA; 216057 MW; 2EE3310DEEB09B9A CRC64;

Query Match      35.9%; Score 968.5; DB 5; Length 2639;
Best Local Similarity 42.6%; Pred. No. 3.2e-41;
Matches 271; Conservative 24; Mismatches 177; Indels 164; Gaps 24;

QY  20 GLGGGAGGQ-----AGAAAAAGGAGGQGGYGLGSGAGRGGGGAGAAAAAGGAGQ 73
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  161 GAGGAGGGGSDSAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 220
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  74 GYGGGLGSGAGRGGLGGGAGAAA---AGVGGQ---GLGGGAGGAGGAGAAAA--- 124
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  221 GYGGYGGYGSAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 280
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  125 ---GGAGGGYGLGSGQ-----GAGRG---GSGGAGAAAA--- 156
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  281 AGSAGGAGGGYGGYGSAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 340
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  157 -----AGGAGGGYGLGSGAG-----RGQ-----LGQAGAAAAAAGGA 194
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  341 AAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 400
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  195 GGGYGLGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 229
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  401 GAGGAGG---GGYGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 458
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  230 ----AGAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 278
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  459 AAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 516
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  279 A-GAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 324
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  517 SDSAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 576
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  325 AAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 361
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  577 AAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 636
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  362 QGAGGAGAAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 420
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  637 YGSDSAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 691
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  421 GGGYGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 473
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  692 GYGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 751
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  474 VSSAVSNLVASPTNSALSSITSNVVSQIGASNPG 509
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  752 HDSAGSAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 778
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
Q8V120 PRELIMINARY; PRT; 1715 AA.
AC Q8V120;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE PE_PGRS family protein.
GN Mt3612.1.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterinene; Mycobacteriaceae; Mycobacterium.
NCBI_TaxID=1773;
RN [1]
```

```

RP  SEQUENCE FROM N.A.
RC  STRAIN=CDC 1551 / OSHKOSH;
RA  Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA  Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA  Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA  Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA  Bishai W.;
RT  "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT  laboratory strains.";
RL  Submitted (Apr-2001) to the EMBL/GenBank/DDBJ databases.
DR  EMBL; AE007163; AKA47971.1; -.
DR  TIGR; MT3612; -.
DR  InterPro; IPR000084; PE_region.
DR  InterPro; IPR002173; PfKB.
DR  Pfam; PF00934; PE; 1.
DR  ProDom; PD001223; PE_region; 1.
DR  PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN_4.
SQ  SEQUENCE 1715 AA; 133832 MW; B857B2A774C771B0 CRC64;

Query Match      34.9%; Score 942.5; DB 16; Length 1715;
Best Local Similarity 37.4%; Pred. No. 1.7e-40;
Matches 267; Conservative 20; Mismatches 224; Indels 203; Gaps 30;

QY  2 ASWTGGQ---QMGRIRIRYGGGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 53
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  497 AAGTGGTGVVGAAGKAGIGGTGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 556
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  54 A-----GRGGGAGAAAAAAGAGGQ-----GGYGLGSGAGRGGLGGGAGGA 96
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  557 GNTGVGGTNGSGGGGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 615
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  97 AA--AAGGV---GGGLGGGAGGAGGAG---AAAAAGGAGGAGGAGGAGGAGGAGG 140
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  616 TATGATGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 675
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  141 GRGGGGGAGAAAAAAGGA---GGYGG---LGSQ-----AGRGGLGQ---GAGAAAAA- 189
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  676 GAGGTGGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 735
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  190 --AAGGAG---GGYGGGLGGGGA-----GQGGYGLGSGAGRGGLGGGAG- 230
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  736 SATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 794
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  231 ---GAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 271
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  795 AGGAGGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 854
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  272 ---AGRGGGAGAAAAAAG---AGGGYGGGGA-----AGGGYGGGGA- 299
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  855 FDGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 914
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  300 -----GQGGYGLGSGQ-----AGRGGLGQ-----GAGAAAAA 328
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  915 NAGVGLTAKAGDGAAGNGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 974
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  329 AAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 386
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  975 ANGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1034
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  387 YGGLGSGQA-----GRGGGAGAAAAAG---AGGGTGGGLGGGQV----- 425
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  1035 VGGDGGEGASGLGLGSLGFDGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1094
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  426 -----GAGGLGGGAGAAAAAAG---GAGGGYGGVSGSASASA 460
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  1095 GCPDGGDGGGGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1154
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  461 AASRLSSPQASSRSVSSAVSNLVASPTNSALSSITSNVVSQIGASNPGSLSCD 514
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  1155 GADGDPIDGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1208
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
```



***This Page Blank (uspto)***



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:05:10 ; Search time 22,7682 Seconds  
(without alignments)  
1995.696 Million cell updates/sec

Title: US-09-490-291-9  
Perfect score: 1809  
Sequence: 1 AEIYNKDGKVDLYKAVGL.....NKLGVGSDTVAIVGYQFA 341

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002:\*

1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1809	100.0	341	22	AA82612
2	1805	99.8	362	20	AA42550
3	1792	99.1	340	17	AAW92998
4	1051	58.1	367	20	AA42549
5	1051	58.1	367	20	AA34058
6	1051	58.1	367	21	AA57356
7	1035	57.2	377	22	AA57356
8	1033	57.1	367	23	ABB08211
9	1029	56.9	377	20	AA34057
10	1029	56.9	377	21	AA57355

11	972	53.7	573	22	ABG28703	Novel human diagno
12	957.5	52.9	328	22	ABG25541	Novel human diagno
13	957.5	52.9	759	22	ABG25868	Novel human diagno
14	957.5	52.9	759	22	ABG29793	Novel human diagno
15	946	52.3	366	23	AB806345	C. bifermentans DP
16	942	52.1	323	23	AAE18784	Sequence of ompK36
17	858.5	47.5	323	20	AAV34056	E. coli outer memb
18	858.5	47.5	323	21	AA57354	E. coli outer memb
19	752	41.6	393	22	ABG25542	Novel human diagno
20	655	36.2	145	22	AAU20470	Human secreted pro
21	623.5	34.5	440	22	ABG15272	Novel human diagno
22	482	26.6	501	22	ABG25865	Novel human diagno
23	377.5	20.9	518	22	ABG18157	Novel human diagno
24	362.5	20.0	122	22	AA25524	Human protein sequ
25	331.5	18.3	1262	22	ABG29003	Novel human diagno
26	276	15.3	89	18	AAW27781	Amino acid sequenc
27	267.5	14.8	702	22	ABG18349	Novel human diagno
28	267.5	14.8	716	22	ABG25880	Novel human diagno
29	267.5	14.8	924	22	ABG25166	Novel human diagno
30	247.5	13.7	353	22	AAE09801	Aeromonas hydrophi
31	247.5	13.7	373	22	AAE09799	Aeromonas hydrophi
32	170.5	9.4	353	20	AA26025	OmpH protein of H.
33	162.5	9.0	353	20	AA26023	OmpH protein of Pa
34	162	9.0	353	11	AAE06038	Class II outer mem
35	157	8.7	342	16	AAE07063	Mature class 2 por
36	157	8.7	342	18	AAW21742	Neisseria meningit
37	157	8.7	363	16	AAW21742	Fusion Class 2 por
38	157	8.7	363	18	AAW21743	Neisseria meningit
39	152.5	8.4	343	20	AA26024	OmpH protein of Pa
40	150.5	8.3	313	18	AAW21744	Neisseria meningit
41	147.5	8.2	309	16	AAW21742	Meningococcal grou
42	147.5	8.2	309	18	AAW21741	Neisseria meningit
43	144.5	8.0	341	11	AAE07043	P3 gene product of
44	140	7.7	372	16	AAE69607	Conococcal porin-5
45	138.5	7.7	342	16	AAE66879	H. Influenzae prot

#### ALIGNMENTS

RESULT 1  
AA82612  
ID AA82612 standard; Protein; 341 AA.  
XX AA82612;  
AC AA82612;  
XX 02-OCT-2001 (first entry)  
XX DT  
XX E. coli outer membrane protein ompF.  
DE  
XX Outer membrane protein; ompF; structural protein; purification.  
XX  
XX Escherichia coli.  
XX OS  
XX WO200153333-A1.  
XX PD  
XX 26-JUL-2001.  
XX  
XX 01-NOV-2000; 2000WO-US30086.  
XX  
XX 20-JAN-2000; 2000US-0490291.  
XX  
PA (MELL/) MELLO C M.  
PA (ARC/) ARCDIACONO S.  
PA (BUTL/) BUTLER M M.  
XX (USSA ) US SEC OF ARMY.  
XX  
PI Mello CM, Arcidiacono S, Butler MM;  
XX WPI; 2001-483136/52.  
XX  
DR Recovering structural polypeptides in a biological sample, useful for  
XX purifying and spinning spider silks and other structural proteins,  
PT

PT comprises treating the sample containing the polypeptides with an acid

XX Claim 2; Page 43-44; 49pp; English.

XX The present sequence is that of the Escherichia coli ompF outer  
XX structural proteins. The invention provides methods for purifying  
XX cells or other biological samples (such as non-recombinant  
XX derived cells), and enrich the purity and yields of structural  
XX proteins by hydrolysing many of the macromolecules while leaving  
XX the structural proteins intact. In the present case, lyophilised  
XX pellets of E. coli were lysed in 9.2 N valeric acid for 1 hr at  
XX room temperature. The cell lysate was clarified by centrifugation  
XX and applied to an SDS polyacrylamide gel for electrophoresis. The  
XX ompF protein was then blotted onto a nitrocellulose membrane for  
XX N-terminal sequencing. The simple extraction procedure yielded  
XX ompF in approximately 8% purity. The new method has the  
XX following advantages over prior art: it involves fewer steps,  
XX requires less time and smaller volumes of reagents, results in  
XX better recovery of protein at higher purity (70-99%), is easy to  
XX scale up, and fibres can be spun in an environmentally benign  
XX solution reducing hazardous waste accumulation and cost.

XX Sequence 341 AA;

Query Match 100.0%; Score 1809; DB 22; Length 341;  
Best Local Similarity 100.0%; Pred. No. 1.1e-147;  
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ABIIYKNDGKNDLYKAVGLHYFSKNGENSYGGNGDMTYARLGFKEGTQINSDLTGYGQ 60  
DB 1 ABIIYKNDGKNDLYKAVGLHYFSKNGENSYGGNGDMTYARLGFKEGTQINSDLTGYGQ 60  
QY 61 WEYNFQGNNSGADATQGNKTRLAFAGLYADVGSDYGRNYYVYDALGYTDMLPFEGG 120  
DB 61 WEYNFQGNNSGADATQGNKTRLAFAGLYADVGSDYGRNYYVYDALGYTDMLPFEGG 120  
QY 121 DTAISDDFFVGRVGVVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGGSISY 180  
DB 121 DTAISDDFFVGRVGVVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGGSISY 180  
QY 181 EYEGFGIYGAYCAADRNLQEAQPLGNGKKAQWATGLKYDANNIYLAANYGETRNPATPI 240  
DB 181 EYEGFGIYGAYCAADRNLQEAQPLGNGKKAQWATGLKYDANNIYLAANYGETRNPATPI 240  
QY 241 TNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYTKSKAKDVEGIGDVLVNYFEVGAT 300  
DB 241 TNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYTKSKAKDVEGIGDVLVNYFEVGAT 300  
QY 301 YYFNKNMSTYVDYIINQIDSDNKLGVGSDDTVAVGIVYQFA 341  
DB 301 YYFNKNMSTYVDYIINQIDSDNKLGVGSDDTVAVGIVYQFA 341

RESULT 2

AA42550  
ID AA42550 standard; Protein; 362 AA.

XX AC AA42550;

XX 20-DEC-1999 (first entry)

XX E. coli wild-type ompF protein.

XX Bacteria; attenuation; deletion; mutant; vaccine; immune response;  
KW Gram negative; infection; diarrhoea; food poisoning; typhoid;  
KW salmonellosis; gonorrhoea; gastroenteritis; whooping cough.

OS Escherichia coli.

PN W09949026-A1.

XX

PD

XX 30-SEP-1999.

PF XX 25-MAR-1999; 99WO-GB00935.

PR XX 25-MAR-1998; 98GB-0006449.

XX (PEPT-) PEPTIDE THERAPEUTICS LTD.  
XX Chatfield SN;

XX WPI; 1999-580447/49.

DR N-PSDB; AA222887.

XX New attenuated bacteria useful as vaccines for protecting against  
PT infections -  
PS Disclosure; Page 63-64; 69pp; English.

XX This sequence represents the E. coli wild-type ompF protein. The coding  
CC sequence of the ompF gene was removed via PCR using primers TT1-TT4  
CC (AA222889-222893) to produce a non-reverting deletion mutation. The  
CC mutant ompF gene (AA222888) was used in the production of a bacterium  
CC attenuated by a non-reverting mutation in each of the ompF gene, the aroC  
CC gene (AA222883, AA222884), and the ompC gene (AA222885, AA222886). The  
CC mutant bacteria provide immunogenic activity with reduced virulence and  
CC thus can be used as a vaccine for raising an immune response against a  
CC variety of bacteria in a mammalian host. Such vaccines can provide  
CC protection against e.g., E. coli (a cause of diarrhoea in humans),  
CC Salmonella typhimurium (the cause of salmonellosis in several animal  
CC species), S. typhi (the cause of human typhoid), S. enteritidis (a cause  
CC of food poisoning in humans), S. choleraesuis (a cause of salmonellosis  
CC in pigs), S. dublin (a cause of both a systemic and diarrhoeal disease in  
CC cattle, especially of new-born calves), Haemophilus influenzae (a cause  
CC of meningitis), Neisseria gonorrhoeae (a cause of gonorrhoeae), Yersinia  
CC enterocolitica (the cause of a spectrum of disease in humans ranging  
CC from gastroenteritis to fatal septicemic disease), Bordetella pertussis  
CC (the cause of whooping cough), and Brucella abortus (a cause of abortion  
CC and infertility in cattle and a condition known as undulant fever in  
CC humans).

XX Query Match 99.8%; Score 1805; DB 20; Length 362;  
Best Local Similarity 100.0%; Pred. No. 2.6e-147;  
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEIYKNDGKNDLYKAVGLHYFSKNGENSYGGNGDMTYARLGFKEGTQINSDLTGYGQ 60  
DB 23 AEIYKNDGKNDLYKAVGLHYFSKNGENSYGGNGDMTYARLGFKEGTQINSDLTGYGQ 82  
QY 61 WEYNFQGNNSGADATQGNKTRLAFAGLYADVGSDYGRNYYVYDALGYTDMLPFEGG 120  
DB 83 WEYNFQGNNSGADATQGNKTRLAFAGLYADVGSDYGRNYYVYDALGYTDMLPFEGG 142  
QY 121 DTAISDDFFVGRVGVVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGGSISY 180  
DB 143 DTAISDDFFVGRVGVVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGGSISY 202  
QY 181 EYEGFGIYGAYCAADRNLQEAQPLGNGKKAQWATGLKYDANNIYLAANYGETRNPATPI 240  
DB 203 EYEGFGIYGAYCAADRNLQEAQPLGNGKKAQWATGLKYDANNIYLAANYGETRNPATPI 262  
QY 241 TNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYTKSKAKDVEGIGDVLVNYFEVGAT 300  
DB 263 TNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYTKSKAKDVEGIGDVLVNYFEVGAT 322  
QY 301 YYFNKNMSTYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 340  
DB 323 YYFNKNMSTYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 362

RESULT 3  
AAW92998

```
ID AAW92998 standard; protein; 340 AA.
XX
AC AAW92998;
XX
DT 19-MAY-1999 (first entry)
XX
DE E. coli OmpF porin protein.
XX
KW OmpF: porin; amphiphilic alpha-structure; beta-structure; GLUT;
KW membrane protein; glucose transporter protein; function.
XX
OS Escherichia coli.
XX
PN WO9618957-A1.
XX
PD 20-JUN-1996.
XX
PF 13-DEC-1995; 95WO-US16126.
XX
PR 14-DEC-1994; 94US-0355844.
XX
PA (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI Cheung M, Czegledy F, Fischbarg J, Iserovich P;
PI Li J;
XX
DR WPI; 1996-300839/30.
XX
PT Predicting tendency to form amphiphilic alpha and beta structures -
PT using a novel algorithm to calculate values for subsequent graphical
PT analysis to predict protein structure
XX
PS Disclosure; Fig 2; 106pp; English.
XX
CC This invention describes a novel method for predicting the tendency of
CC a protein to form either: (a) an amphiphilic alpha-structure or (b) an
CC amphiphilic beta-structure. The methods are used particularly for
CC predicting the structure of membrane proteins such as glucose transporter
CC proteins (GLUTs). They can be used to discern the function of proteins.
CC They can also be used for the rational design or identification of
CC compounds which interact with the proteins or to engineer proteins having
CC particular structures. This sequence represents an Escherichia coli
CC OmpF porin which is used to illustrate the method of the invention.
XX
SQ Sequence 340 AA;

Query Match 99.1%; Score 1792; DB 17; Length 340;
Best Local Similarity 99.4%; Pred. No. 3.2e-146;
Matches 338; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AEIYNKDNKVDLYGRAVGLHYFSKNGENSGNGDMTYARLGFKGETQINSDLTGYGQ 60
DB 1 AEIYNKDNKVDLYGRAVGLHYFSKNGENSGNGDMTYARLGFKGETQINSDLTGYGQ 60
QY 61 WEYNFQGNNEGADAOQGNKTRLAFAAGLYADVGSFDRYGRNYGVVYDALGYTDLMPFEGG 120
DB 61 WEYNFQGNNEGADAOQGNKTRLAFAAGLYADVGSFDRYGRNYGVVYDALGYTDLMPFEGG 120
QY 121 DTAYSDDFFVGRVGGVATYRNSNFFGLVDGLNFAVOYLGNKNERDTARRSNGDVGGSISY 180
DB 121 DTAYSDDFFVGRVGGVATYRNSNFFGLVDGLNFAVOYLGNKNERDTARRSNGDVGGSISY 180
QY 181 EYEGFGIVGAYGAADRNLQEAQPLNGKKAQWATGLKDYDANNIYLAANYGETRATPI 240
DB 181 EYEGFGIVGAYGAADRNLQEAQPLNGKKAQWATGLKDYDANNIYLAANYGETRATPI 240
QY 241 TNKFTNTSGFANKTQDVLVLAQYQFFGLRPSIATYTKSKAKDVEGIGDVLVNYFEVGAT 300
DB 241 TNKFTNTSGFANKTQDVLVLAQYQFFGLRPSIATYTKSKAKDVEGIGDVLVNYFEVGAT 300
QY 301 YFYNKNMSTVYDIINQIDSDNKLGVGSDDTVAAGIVYQF 340
DB 301 YFYNKNMSTVYDIINQIDSDNKLGVGSDDTVAAGIVYQF 340
```

---

```
RESULT 4
AAI42549
ID AAY42549 standard; Protein; 367 AA.
XX
AC AAY42549;
XX
DT 20-DEC-1999 (first entry)
XX
DE E. coli wild-type ompC protein.
XX
KW Bacteria: attenuation; deletion; mutant; vaccine; immune response;
KW Gram negative; infection; diarrhoea; food poisoning; typhoid;
KW salmonellosis; gonorrhoea; gastroenteritis; whooping cough.
XX
OS Escherichia coli.
XX
PN WO9949026-A1.
XX
PD 30-SEP-1999.
XX
PF 25-MAR-1999; 99WO-GB00935.
XX
PR 25-MAR-1998; 98GB-0006449.
XX
PA (PEPT-) PEPTIDE THERAPEUTICS LTD.
XX
PI Chatfield SN;
XX
DR WPI; 1999-580447/49.
DR N-PSDB; AA222885.
XX
PT New attenuated bacteria useful as vaccines for protecting against
PT infections -
XX
PS Disclosure; Page 57-58; 69pp; English.
XX
CC This sequence represents the E. coli wild-type ompC protein. The coding
CC sequence of the ompC gene was removed via PCR using primers TT7-TT10
CC (AA222883-422886) to produce a non-reverting deletion mutation. The
CC mutant ompC gene (AA222886) was used in the production of a bacterium
CC attenuated by a non-reverting mutation in each of the ompC gene, the aroC
CC gene (AA222883, AA222884), and the ompF gene (AA222887, AA222888). The
CC mutant bacteria provide immunogenic activity with reduced virulence and
CC variety of bacteria in a mammalian host. Such vaccines can provide a
CC protection against e.g., E. coli (a cause of diarrhoea in humans),
CC Salmonella typhimurium (the cause of salmonellosis in several animal
CC species), S. typhi (the cause of human typhoid), S. enteritidis (a cause
CC of food poisoning in humans), S. choleraesuis (a cause of salmonellosis
CC in pigs), S. dublin (a cause of both a systemic and diarrhoeal disease in
CC cattle, especially of new-born calves), Haemophilus influenzae (a cause
CC of meningitis), Neisseria gonorrhoeae (a cause of gonorrhoea), Yersinia
CC enterocolitica (the cause of a spectrum of disease in humans ranging
CC from gastroenteritis to fatal septicaemic disease), Bordetella pertussis
CC (the cause of whooping cough), and Brucella abortus (a cause of abortion
CC and infertility in cattle and a condition known as undulant fever in
CC humans).
XX
SQ Sequence 367 AA;

Query Match 58.1%; Score 1051; DB 20; Length 367;
Best Local Similarity 58.9%; Pred. No. 2.5e-82;
Matches 212; Conservative 45; Mismatches 69; Indels 34; Gaps 8;

QY 1 AEIYNKDNKVDLYGRAVGLHYFSKNGENSGNGDMTYARLGFKGETQINSDLTGYGQ 60
DB 22 AEIYNKDNKVDLYGRAVGLHYFSKNGENSGNGDMTYARLGFKGETQINSDLTGYGQ 76
QY 61 WEYNFQGNNEGADAOQGNKTRLAFAAGLYADVGSFDRYGRNYGVVYDALGYTDLMPFEGG 120
DB 77 WEYQIQGNSAEN---ENNSWTRVAFAGLKFDQVGSFDRYGRNYGVVYDTSVDTVLPEFG 133
```



```

XX SQ Sequence 367 AA;
Query Match 58.1%; Score 1051; DB 21; Length 367;
Best Local Similarity 58.9%; Pred. No. 2.5e-82;
Matches 212; Conservative 45; Mismatches 69; Indels 34; Gaps 8;

QY 1 AEIYNKDGKVDLYGKAVGLHYFSKNGENSYGCGDMTYARLGFKGETQINSLDTGYGQ 60
DB 22 AEVYNKDGKNDLYGKVDGLHYFS-----DNKVDGDTYMRLGFKGETQTDLTGYGQ 76
QY 61 WEYNFQGNNSGADAQGTGNKTRLAFAGLKADYGVSDYGRNYGVVYDALGYTDMLEPFGG 120
DB 77 WEYQIQGNSAEN---ENNSTWTFVAFAGLKFDQVGSFSDYGRNYGVVYDVTWTDVLEPFGG 133
QY 121 DTAYSDDFFVGRVGGVATYRNSNFFGLVDGLNFAVQYLGKNERDT-----A 166
DB 134 DTYSDDNFMQGRNGGFATYRTDFFGLVDGLNFAVQYQKNGNSPGEFTSGVTNNGRDA 193
QY 167 RRSNGDVGGSISYEYEGFGIVGAYGAADRTNLOE-AOPLGNGKKAQOWATGLKAYDANNI 225
DB 194 LRQNGDVGGSITDYEGFGIGGAISSKRTDAQNTAAAYICNGDRAETTYGGLAYDANNI 253
QY 226 YLAANYGETRNATPTNKFTNTSGFANKTDVLLVAQYQDFGLRPSIATYTKSKADY-E 284
DB 254 YLAAYQTQYNATRV-----GSLGWANKAQNEFAVQYQDFGLRPSIAYLQSKGNLGR 308
QY 285 GIGDVLNTEFVCGATYFNKNMSTYVDYIINQIDSNKL-----GVGSDDTVAVGIYQF 340
DB 309 GYDEDDILKYVDGATYFNKNMSTYVDYKINLLD-DNQFTRDAGINTDNIVALGLVYQF 367

RESULT 7
AAG98980
ID AAG98980 standard; Protein; 377 AA.
XX AC AAG98980;
XX DT 26-SEP-2001 (first entry)
XX DE E. coli growth and proliferation related protein sequence SEQ ID NO:450.
XX KW Escherichia coli; growth; proliferation; microbial; antimicrobial;
XX KW bacterial infection; microorganism.
XX OS Escherichia coli.
XX PN WO200134810-A2.
XX PD 17-MAY-2001.
XX PF 09-NOV-2000; 2000WO-US30950.
XX PR 09-NOV-1999; 99US-0164415.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Forsyth RA, Ohlsen K, Zyskind J;
XX DR WPI; 2001-335933/35.
XX DR N-PSDB; AAH84651.
XX PT Novel nucleic acids that inhibit Escherichia coli proliferation, useful
XX PT for screening for homologous genes and for designing expression vectors
XX PS Claim 19; Page 505-506; 522pp; English.
XX CC AAH84373 to AAH84499 represent Escherichia coli growth and proliferation
XX CC related DNA sequences (I). AAH84500 to AAH84670 encode the E. coli
XX CC growth and proliferation related proteins given in AAG99078 and AAG98830
XX CC to AAG98999. (I) can be used as potential targets for the generation of
XX CC new antimicrobial agents, and for identification of compounds which

```

interact with the gene products of (I). In addition the expression of (I) and the purification of the proteins, the purified proteins can be used to generate reagents and screen small molecule libraries or other candidate compound libraries for compounds that can be further developed to yield novel antimicrobial compounds. In addition, nucleic acid probes complementary to (I) that are specific for particular species of microorganisms can be used to identify particular microorganism species in clinical specimens, therefore, providing a rapid and dependable method by which to identify the causative agents of a bacterial infection. Also, antibodies generated against proteins translated from mRNA transcribed from proliferation-required sequences can also be used to screen for specific microorganisms that produce such proteins in a species-specific manner. AAH84371 and AAH84670 represent sequencing primers used in the isolation of E. coli growth and proliferation related sequence, which are used in an example from the present invention.

Query Match 57.2%; Score 1035; DB 22; Length 377;  
Best Local Similarity 56.9%; Pred. No. 6.3e-81;  
Matches 209; Conservative 42; Mismatches 78; Indels 38; Gaps 9;

QY 1 AEIYNKDGKVDLYGKAVGLHYFSKNGENSYGCGDMTYARLGFKGETQINSLDTGYGQ 60  
DB 22 AEVYNKDGKNDLYGKVDGLHYFSDSNAK-----DGDSYARLGFKGETQINDLTGYGQ 76  
QY 61 WEYNFQGNNSGADAQGTGNKTRLAFAGLKADYGVSDYGRNYGVVYDALGYTDMLEPFGG 120  
DB 77 WEYNIQANTESSKNQSW--TRLAFAGLKAFADYGVSDYGRNYGVYDIEGTMDLPEFGG 134  
QY 121 DT-AYSDDFFVGRVGGVATYRNSNFFGLVDGLNFAVQYLGKNE-----RDTAR 167  
DB 135 DSYTNADNFMGTGRANGVATYRTDFFGLVNLNFAVQYQKNGNSQEGTNNGRD-VR 193  
QY 168 RNSGDVGGSISYEY-EGFGIVGAYGAADRTNLOEAOPLGNGKKAQOWATGLKYDANNIY 226  
DB 194 HENGDGWGLSTYDILGWFSGAAGAYTSSDRTNDQVNHNTAAGDKADAWTAGLKVDANNIY 253  
QY 227 LAANYGETRNATPTNKFTNTSGFANKTDVLLVAQYQDFGLRPSIATYTKSKADVE-- 284  
DB 254 LATYSETRNTPTFGD---SDYAVANKTQNFVTAQYQDFGLRPAVSFLMSKGRDLHA 310  
QY 285 -----GIGDVLNTEFVCGATYFNKNMSTYVDYIINQIDSNKL-----GVGSDDTVA 333  
DB 311 GGADNPAGVDDKOLVKYADIGATYFNKNMSTYVDYKINLLDEDDSFYAANGISTDDIVA 370  
QY 334 VGIVYQF 340  
DB 371 LGLVYQF 377

RESULT 8  
ABB08211  
ID ABB08211 standard; protein; 367 AA.  
XX AC ABB08211;  
XX DT 08-APR-2002 (first entry)  
XX DE Escherichia coli outer membrane protein C (OmpC) precursor.  
XX KW OmpC; outer membrane protein C; Crohn's disease; immunoglobulin A; IgA;  
XX KW antiinflammatory; immune response.  
XX OS Escherichia coli.  
XX PN WO200189361-A2.  
XX PD 29-NOV-2001.  
XX PF 17-MAY-2001; 2001WO-US16032.

```

PR 19-MAY-2000; 2000US-0575061.
PA (CEDA-) CEDARS SINAI MEDICAL CENT.
PA (REGC ) UNIV CALIFORNIA.
PI Targan SR, Braun J, Sutton CL;
XX WPI; 2002-075333/10.
XX
XX Diagnosing Crohn's disease in a subject comprises detecting the
XX presence of immunoglobulin A anti-outer membrane protein C antibodies -
XX Claim 3; Fig 5; 55pp; English.
XX
XX The sequence represents the E. coli outer membrane protein C (OmpC)
XX precursor. The invention relates to a novel method for diagnosing Crohn's
XX disease in a subject. The method comprises detecting immunoglobulin (Ig)A
XX OmpC antibodies in the subject, where the presence of the IgA anti-OmpC
XX antibodies indicates that the subject has Crohn's disease. The proteins
XX of the invention have antiinflammatory activity. The method of the
XX invention is an immunological response elicitor. The method is useful for
XX diagnosing Crohn's disease in a subject and administering an OmpC antigen
XX is useful for inducing tolerance in a patient with Crohn's disease.
XX
XX Sequence 367 AA;
XX
XX Query Match 57.1%; Score 1033; DB 23; Length 367;
XX Best Local Similarity 58.3%; Pred. No. 9e-81;
XX Matches 210; Conservative 45; Mismatches 71; Indels 34; Gaps 8;
XX
XX QY 1 AEIYNKDGKNDLYGKAVGLHYFSKNGNSYGGNDMTYARLGFKGETQINSDLTGYYGO 60
XX DB 22 AEIYNKDGKNDLYGKAVGLHYFSKNGNSYGGNDMTYARLGFKGETQINSDLTGYYGO 76
XX
XX QY 61 WEYNFOGNNSEGADAOQGNKTRLAFAGLVADYSGDFYGRNRYGVVYDALGYTDMLEPFGG 120
XX DB 77 WEYQIQGNSAEN---ENNNTWRVAFAGLVADYSGDFYGRNRYGVVYDALGYTDMLEPFGG 133
XX
XX QY 121 DTAYSDDFVGVGVVATYRNNSFFGLVDGLNFAVQYLGKNERDT-----A 166
XX DB 134 DTYGSDNFNQQRNGFATYRNNTDFGLVDGLNFAVQYLGKNGNPSGEGFTSGVTNNGRDA 193
XX
XX QY 167 RNSGNGVCGSISYEYEGFIVGAYGAADRTNLQE-AQPLGNGKKAQWATGLKYDANNI 225
XX DB 194 LRQNGDVGGSITDYEGFGIGGATSSSRRTDAQNTAAIYGNDRAEATYTGGLKYDANNI 253
XX
XX QY 226 YLAANYGETRNATPITNKFTNTSGFANKTQDVLVVAQYQDFGLRPSIATYTSKAKDV-E 284
XX DB 254 YLAANYQTYNATRV-----GSLGWANKAQNFVAQYQDFGLRPSLAYLQSKGNLGR 308
XX
XX QY 285 GIGDVLVNYFEVGATYTFYFNKNMSTYVDYIINQIDSNDKL-----GVGSDDTVAVGIVYQF 340
XX DB 309 GYDDEDILKYVDVGATYTFYFNKNMSTYVDYIINQIDSNDKL-DNQFTRDAGINTDNIVALGLVYQF 367
XX
XX RESULT 9
XX AAY34057
XX ID AAY34057 standard; protein; 377 AA.
XX AC AAY34057;
XX
XX XX 23-NOV-1999 (first entry)
XX
XX DE E. coli outer membrane protein F precursor.
XX
XX KW Ulcerative colitis; histone; H1-like antigen; porin antigen;
XX KW Bacteroides antigen; IB; PANCA; inflammatory bowel disease;
XX KW diagnosis; perinuclear anti-neutrophil cytoplasmic antibody;
XX KW outer membrane protein F precursor.
XX
XX OS Escherichia coli.
XX
XX XX W09945955-A1.
XX

```

```

XX 16-SEP-1999.
XX
XX PF 12-MAR-1999; 99WO-US03492.
XX
XX PR 12-MAR-1998; 98US-0041889.
XX
XX PA (REGC ) UNIV CALIFORNIA.
XX
XX PI Braun J, Cohavy O;
XX WPI; 1999-551215/46.
XX
XX Use of histone H1, porin or Bacteroides antigens as targets for the
XX diagnosis, prevention and treatment of ulcerative colitis -
XX Claim 8; Fig 11; 134pp; English.
XX
XX The invention provides a method for the diagnosis, prevention and
XX treatment of ulcerative colitis (UC) using histone H1-like antigen, a
XX porin antigen or a Bacteroides antigen as a target antigen. The novel
XX method of diagnosing UC in a subject suspected of having inflammatory
XX bowel disease (IBD) comprises: (1) obtaining a sample from the subject;
XX (2) contacting the sample with a histone H1-like antigen, or perinuclear
XX anti-neutrophil cytoplasmic antibody (pANCA)-reactive fragment, to form a
XX complex of the histone H1-like antigen, or the PANCA-reactive fragment,
XX and antibody to the histone H1-like antigen; and (3) detecting the
XX presence or absence of the complex; where the presence of the complex
XX indicates that the subject has UC. The PANCA-reactive histone H1-like
XX antigen, porin antigen and Bacteroides antigen are useful in the
XX diagnosis, prevention and treatment of UC. The methods can also be used
XX for identifying agents useful for treating UC. The present sequence
XX represents a E. coli outer membrane protein F precursor.
XX
XX SQ Sequence 377 AA;
XX
XX Query Match 56.9%; Score 1029; DB 20; Length 377;
XX Best Local Similarity 56.9%; Pred. No. 2.1e-80;
XX Matches 209; Conservative 41; Mismatches 79; Indels 38; Gaps 9;
XX
XX QY 1 AEIYNKDGKNDLYGKAVGLHYFSKNGNSYGGNDMTYARLGFKGETQINSDLTGYYGO 60
XX DB 22 AEIYNKDGKNDLYGKAVGLHYFSKNGNSYGGNDMTYARLGFKGETQINSDLTGYYGO 76
XX
XX QY 61 WEYNFOGNNSEGADAOQGNKTRLAFAGLVADYSGDFYGRNRYGVVYDALGYTDMLEPFGG 120
XX DB 77 WEYNQIANTTESKNSQW--TRLAFAGLVADYSGDFYGRNRYGVVYDALGYTDMLEPFGG 134
XX
XX QY 121 DT-AYSDDFVGVGVVATYRNNSFFGLVDGLNFAVQYLGKNE-----RDTAR 167
XX DB 135 DSYTNADNFMTCRANGVATYRNNTDFGLVDGLNFAVQYQGNNEGASNGQEGTNNGRD-VR 193
XX
XX QY 168 RNSGNGVCGSISYEY-EGFIVGAYGAADRTNLQEAQPLGNGKKAQWATGLKYDANNI 226
XX DB 194 HENGDGWGLSTTYDLMGMFSGAGAAATSSDRTNDQVNHNTAAGDKADAWTAGLKYDANNI 253
XX
XX QY 227 LAANYGETRNATPITNKFTNTSGFANKTQDVLVVAQYQDFGLRPSIATYTSKAKDVE-- 284
XX DB 254 LATMYSETRNMTFFGD---SDYAVANKTONFEVTAQYQDFGLRPAVSLFMSKGRDLHAA 310
XX
XX QY 285 -----GIGDVLVNYFEVGATYTFYFNKNMSTYVDYIINQIDSNDKL-----GVGSDDTVA 333
XX DB 311 GGADNPAGVDDKDLVKYADIGATYTFYFNKNMSTYVDYIINQIDSNDKLDEDDSFYAANGISTDDIVA 370
XX
XX QY 334 VGIVYQF 340
XX DB 371 LGLVYQF 377
XX
XX RESULT 10
XX AAY57355
XX ID AAY57355 standard; protein; 377 AA.
XX

```

AA57355;  
 13-JUN-2000 (first entry)  
 E. coli outer membrane protein F precursor.  
 Ulcerative colitis; inflammatory bowel disease; porin antigen; MAb;  
 pANCA; perinuclear anti-neutrophil cytoplasmic antibody;  
 histone H1; outer membrane protein F precursor.  
 Escherichia coli.  
 US6033864-A.  
 07-MAR-2000.  
 12-MAR-1998; 98US-0041889.  
 12-APR-1996; 96US-0057846.  
 11-APR-1997; 97US-0837058.  
 (REGC ) UNIV CALIFORNIA.  
 Cohavy O, Braun J;  
 WPI; 2000-255695/22.  
 Diagnosing ulcerative colitis or susceptibility, by detecting complex  
 formation between microbial porin antigen and perinuclear  
 anti-neutrophil cytoplasmic autoantibodies -  
 Claim 1; Fig 10; 49pp; English.  
 The invention provides a method for diagnosing ulcerative colitis in a  
 subject suspected of having inflammatory bowel disease. The method  
 comprises reacting a patient sample with a porin antigen that is  
 immunologically reactive with pANCA (perinuclear anti-neutrophil  
 cytoplasmic antibodies) and detecting formation of a Ag-PANCA complex  
 as indicative of ulcerative colitis. The method is used to diagnose  
 ulcerative colitis or susceptibility to it. The present sequence  
 represents a E. coli outer membrane protein F precursor.  
 Claim 1; Fig 10; 49pp; English.  
 Query Match 56.9%; Score 1029; DB 21; Length 377;  
 Best Local Similarity 56.9%; Pred. No. 2.1e-80;  
 Matches 209; Conservative 41; Mismatches 79; Indels 38; Gaps 9;  
 1 AEIYKDKGNKVDLYGRAVGLHYFSGKNGENSYGGNGDMTYARLFGKGTQINSDLTGYGQ 60  
 22 AEIYKDKGNKLDLYGKVDGLHYFSDNSAK-----DGDQSYARLFGKGTQINDLTGYQQ 76  
 61 WEYNFGNNSGADAGQNGKTRLAFLAGLKYADYVGSFEDYGRNYGVVYDALGYTDMLPFEGG 120  
 77 WEYNIQANTNESSKQNSW--TRLAFLAGLKFADYGSFEDYGRNYGVYDIEGWTMDLPFEGG 134  
 121 DT-AYSDFFVGRVGGVATYRNSNFCLVDGLNFAVQYLCKNE-----RRTAR 167  
 135 DSTNADNFMTGRANGVATYRNTDFGLVNLNFAVQYQGNNEGASNGQEGTNNGRD-VR 193  
 168 RNSGDGVGSGISVEY-EGFIVGAYGAADRNLQEAQPLNGKKAQOWATGLKYDANNIY 226  
 194 HENGDWGLSTTYDLGNGFFAGAYYSSTDRNTDQVNHNTAAGGKADAWTAGLYKDYANNIY 253  
 227 LAANYGSTRNATPITNKTNTSGFANKTQDVLVLAQYQDFGLRPSIATYKSKAKDVE-- 284  
 254 LATMYSETRNTMPFGD---SDYAVANKTQNEFVTAQYQDFGLRPAVSFLMSKGRDLHAA 310  
 285 -----GIGDVLVYFEVGATYFFKNMSTYVDYIINOIDSNDKL----GVGSDDTVA 333  
 311 GGADNPAGVDKDLVRYADYIGATYFFKNMSTYVDYKINLIDEDDSFYAANGISTDDIVA 370  
 334 VGIYVQF 340

Db 371 LGLVYQF 377  
 RESULT 11  
 ABG28703  
 ID ABG28703 standard; Protein; 573 AA.  
 XX  
 AC ABG28703;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #28694.  
 XX  
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 KW  
 XX Homo sapiens.  
 OS  
 XX WO200175067-A2.  
 PN  
 XX 11-OCT-2001.  
 PD  
 XX 30-MAR-2001; 2001WO-US08631.  
 XX  
 PF 31-MAR-2000; 2000US-0540217.  
 PR  
 XX 23-AUG-2000; 2000US-0649167.  
 PR  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX Drmanac RT, Liu C, Tang YT;  
 PI  
 XX WPI; 2001-639362/73.  
 DR  
 DR N-PSDB; AAS92890.  
 XX  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 XX Claim 20; SEQ ID No 59062; 103pp; English.  
 PS  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 573 AA;  
 Query Match 53.7%; Score 972; DB 22; Length 573;  
 Best Local Similarity 61.0%; Pred. No. 3e-75;  
 Matches 188; Conservative 33; Mismatches 71; Indels 16; Gaps 5;  
 1 AEIYKDKGNKVDLYGRAVGLHYFSGKNGENSYGGNGDMTYARLFGKGTQINSDLTGYGQ 60  
 28 AEIYKDKGNKLDLYGKVDGLHYFSDNSAK-----DNASKDGQSYIRFGKGTQINDLTGYGR 82







Db 715 ATYYFNKNSTYVDYKINLLD-DNQFTRDAGINTDNIVALGLVYQF 759

## RESULT 15

ABB06345

ID ABB06345 standard; Protein; 366 AA.

XX ABB06345;

XX 30-MAY-2002 (first entry)

XX C. bifermentans DPH-1 perchloroethylene dehalogenase SEQ ID NO:1.

XX Clostridium bifermentans DPH-1; perchloroethylene dehalogenase;

KW enzyme; decomposing; chlorinated aliphatic compound.

XX Clostridium bifermentans.

XX JP2002017358-A.

XX 22-JAN-2002.

XX 04-JUL-2000; 2000JP-0202729.

XX 04-JUL-2000; 2000JP-0202729.

XX (GIFU-) GIFU DAIGAKUCHO.

XX WPI; 2002-263241/31.

DR N-PSDB; ABL49767.

XX A perchloroethylene dehalogenase derived from Clostridium bifermentans  
PT DPH-1, for decomposing chlorinated aliphatic compounds. -

XX Claim 4; Page 11-12; 15pp; Japanese.

XX The present sequence represents perchloroethylene (PCE) dehalogenase  
CC derived from Clostridium bifermentans DPH-1. PCE dehalogenase:  
CC (a) catalyses the dehalogenising reaction forming trichloroethylene  
CC from tetrachloroethylene; (b) is a dimer consisting of two subunits  
CC having a molecular weight of 35 kDa; (c) has an optimum temperature  
CC of 35 plus degrees Celsius; (d) has an optimum pH of 7.5; (e) has  
CC enzymatic activity that is not affected by the addition of metal  
CC ions, ethylene diamine tetraacetate, reduced nicotinamide adenine  
CC dinucleotide or cyanocobalamin; and (f) has enzymatic activity which  
CC is inhibited by the presence of oxygen. The enzyme is used for  
CC decomposing chlorinated aliphatic compounds.

SQ Sequence 366 AA;

Query Match 52.3%; Score 946; DB 23; Length 366;

Best Local Similarity 55.8%; Pred. No. 2.8e-73;

Matches 191; Conservative 36; Mismatches 85; Indels 30; Gaps 8;

Qy 1 AEIYNKDGKVDLYKAVGLHYFESKNGENSGYGGNDMTYARLGFKEGTQINSDLTGYGQ 60

Db 22 AEVYNKDGKVDLYKAVGLHYFESKNGENSGYGGNDMTYARLGFKEGTQINSDLTGYGQ 76

Qy 61 WEYNFQGNSEGAQAQTKGNKTRLAFAGLYADVGSDYGRNRYVYDALGYTDMLPFEGG 120

Db 77 WEYQVQANGTEGDKCDSW--TRLAFAGIKVGDYGSFDYGRNRYVYDALGYTDMLPFEGG 134

Qy 121 DT-AYSDDFVGRVGGVATYRNSNFFGLVDGLNFAVQYLGNKNERDT----- 165

Db 135 DSYTKADNFMTRANGVATYRNTDFGLVDGLNVALQYOGANENQSPQEGETNNGGDDRN 194

Qy 166 ARRSNGDVGGSISVEYEGGIV--GAYCAADRTNLQAPLGNCKKAEQWATGLKYDAN 223

Db 195 MKNNGDVGGSISSTYDL-GMGVSTFGAAYTSSDRTNEQYNDSTAGGDKADAWTVGLKYDAN 253

Qy 224 NIYLAANYETERNATPITNKNTNTSGFANKTQDVLVVAQYQDFGLRPSIATYKSKAKDV 283

Db 254 NIYLATWYSETRNTPYGGSGNSDNTIANKNTQNTFEVTAQYQDFGLRPSVFLSKGKDL 313

Qy 284 EGI----GDVLDVNYFEVGATYFKNKMSYVDYIINQIDSD 321

Db 314 -GVNGSDGDQDLVKYASVGATYFKNKMSYVDYIINQIDSD 354

Search completed: December 18, 2002, 16:17:26

Job time : 24.7682 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:14:14 : Search time 7.85111 Seconds  
(without alignments)  
1277.936 Million cell updates/sec

Title: US-09-490-291-9  
Perfect score: 1809  
Sequence: 1 AEIYNKDGKVDLYGKAVGL.....NKLGVGSDDTVAIVGYQFA 341

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA:\*  
1: /cgn2.6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2.6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2.6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2.6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2.6/ptodata/1/iaa/PTUS\_COMB.pep.\*  
6: /cgn2.6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1792	99.1	340	2	US-08-355-844-1
2	1792	99.1	340	5	PCT-US95-16126-1
3	1051	58.1	367	3	US-09-041-889-30
4	1029	56.9	377	3	US-09-041-889-29
5	858.5	47.5	323	3	US-09-041-889-28
6	157	8.7	342	1	US-08-096-182A-4
7	157	8.7	342	1	US-08-877-109-4
8	157	8.7	342	3	PCT-US94-08327-4
9	157	8.7	342	3	US-08-798-760-4
10	157	8.7	363	1	US-08-096-182A-6
11	157	8.7	363	1	US-08-877-109-6
12	157	8.7	363	3	PCT-US94-08327-6
13	157	8.7	363	3	US-08-798-760-6
14	147.5	8.2	309	1	US-08-096-182A-2
15	147.5	8.2	309	1	US-08-877-109-2
16	147.5	8.2	309	3	US-08-798-760-2
17	147.5	8.2	309	5	PCT-US94-08327-2
18	138.5	7.7	342	4	US-08-096-181A-12
19	138.5	7.7	342	4	US-08-096-181A-14
20	138.5	7.7	342	5	PCT-US94-08326-12
21	138.5	7.7	342	5	PCT-US94-08326-14
22	138.5	7.7	361	4	US-08-096-181A-8
23	138.5	7.7	361	5	PCT-US94-08326-8
24	138.5	7.7	363	4	US-08-096-181A-10
25	138.5	7.7	363	5	PCT-US94-08326-10
26	135.5	7.5	350	4	US-09-333-611-14
27	120.5	6.7	326	4	US-09-333-611-12

ALIGNMENTS

RESULT 1

US-08-355-844-1

; Sequence 1, Application US/08355844

; Patent No. 5940307

; GENERAL INFORMATION:

; APPLICANT: Fischbarg, Jorge

; APPLICANT: Czegledy, Ferenc

; APPLICANT: Iserovich, Pavel

; APPLICANT: Li, Jun

; APPLICANT: Cheung, Min

; TITLE OF INVENTION: A METHOD FOR PREDICTING PROTEIN

; TITLE OF INVENTION: STRUCTURE

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond

; STREET: 30 Rockefeller Plaza

; CITY: New York

; STATE: NY

; COUNTRY: USA

; ZIP: 10112-0228

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/355,844

; FILING DATE: 14-DEC-1994

; CLASSIFICATION: 436

; ATTORNEY/AGENT INFORMATION:

; NAME: Tang, Henry Y. S.

; REGISTRATION NUMBER: 29,705

; REFERENCE/DOCKET NUMBER: A29927-50/29910

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-408-2586

; TELEFAX: 212-765-2519

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 340 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; ORIGINAL SOURCE:

; ORGANISM: Escherichia coli

; FEATURE:

; NAME/KEY: Peptide

; LOCATION: 1..340

; OTHER INFORMATION: OmpF porin protein

US-08-355-844-1

```
Query Match          99.1%; Score 1792; DB 2; Length 340;
Best Local Similarity 99.4%; Pred. No. 3e-159;
Matches 338; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AEIYNKDGKVDLYKAVGLHYFSKNGNSYGGNGDMTYARLGFKEGTQINSDLTGYGQ 60
DB 1 AEIYNKDGKVDLYKAVGLHYFSKNGNSYGGNGDMTYARLGFKEGTQINSDLTGYGQ 60
QY 61 WEYNFQGNSEGADQATGKNTLAFAGLYADVGSDYGRNYSYVVDALGYTDMLEPFGG 120
DB 61 WEYNFQGNSEGADQATGKNTLAFAGLYADVGSDYGRNYSYVVDALGYTDMLEPFGG 120
QY 121 DTAYSDDFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGSGSISY 180
DB 121 DTAYSDDFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGSGSISY 180
QY 181 EYEGFGIYGAYGAADRTNLQEAQPLGNGKKAQOWATGLKYDANNIYLAANYGETRNATPI 240
DB 181 EYEGFGIYGAYGAADRTNLQEAQPLGNGKKAQOWATGLKYDANNIYLAANYGETRNATPI 240
QY 241 TNKFTNTSGFANKTDVLLVAQYQDFGLRPSIAIYTKSKAKDVEGIGDVLVNYFEVGAT 300
DB 241 TNKFTNTSGFANKTDVLLVAQYQDFGLRPSIAIYTKSKAKDVEGIGDVLVNYFEVGAT 300
QY 301 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 340
DB 301 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 340

RESULT 2
PCT-US95-16126-1
; Sequence 1, Application PC/TUS9516126
; GENERAL INFORMATION:
; APPLICANT: Fischberg, Jorge
; APPLICANT: Czegledy, Ferenc
; APPLICANT: Iserovich, Pavel
; APPLICANT: Li, Jun
; APPLICANT: Cheung, Min
; TITLE OF INVENTION: A METHOD FOR PREDICTING PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10112-0228
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16126
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/355,844
; FILING DATE: 14-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Tang, Henry Y.S.
; REGISTRATION NUMBER: 29,705
; REFERENCE/DOCKET NUMBER: A29927-50/29910
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-408-2586
; TELEFAX: 212-765-2519
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 340 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
```

```
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..340
; OTHER INFORMATION: OmpF porin protein
PCT-US95-16126-1

Query Match          99.1%; Score 1792; DB 5; Length 340;
Best Local Similarity 99.4%; Pred. No. 3e-159;
Matches 338; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AEIYNKDGKVDLYKAVGLHYFSKNGNSYGGNGDMTYARLGFKEGTQINSDLTGYGQ 60
DB 1 AEIYNKDGKVDLYKAVGLHYFSKNGNSYGGNGDMTYARLGFKEGTQINSDLTGYGQ 60
QY 61 WEYNFQGNSEGADQATGKNTLAFAGLYADVGSDYGRNYSYVVDALGYTDMLEPFGG 120
DB 61 WEYNFQGNSEGADQATGKNTLAFAGLYADVGSDYGRNYSYVVDALGYTDMLEPFGG 120
QY 121 DTAYSDDFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGSGSISY 180
DB 121 DTAYSDDFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGSGSISY 180
QY 181 EYEGFGIYGAYGAADRTNLQEAQPLGNGKKAQOWATGLKYDANNIYLAANYGETRNATPI 240
DB 181 EYEGFGIYGAYGAADRTNLQEAQPLGNGKKAQOWATGLKYDANNIYLAANYGETRNATPI 240
QY 241 TNKFTNTSGFANKTDVLLVAQYQDFGLRPSIAIYTKSKAKDVEGIGDVLVNYFEVGAT 300
DB 241 TNKFTNTSGFANKTDVLLVAQYQDFGLRPSIAIYTKSKAKDVEGIGDVLVNYFEVGAT 300
QY 301 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 340
DB 301 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 340

RESULT 3
US-09-041-889-30
; Sequence 30, Application US/09041889
; Patent No. 6033864
; GENERAL INFORMATION:
; APPLICANT: Braun, Jonathan
; APPLICANT: Cohavy, Offer
; TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
; TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes Thereof, Using
; TITLE OF INVENTION: Microbial UC PANCA antigens
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,889
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/837,058
; FILING DATE: 11-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-PM 3006
```

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 367 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-041-889-30

Query Match 58.1%; Score 1051; DB 3; Length 367;  
Best Local Similarity 58.9%; Pred. No. 4.1e-90;  
Matches 212; Conservative 45; Mismatches 69; Indels 34; Gaps 8;  
QY 1 AEYNDGKNDVLYGKAVGLHYFSKNGENSYGNGDMTYARLFGKGTQINSDLTGYYQ 60  
DB 22 AEYNDGKNDVLYGKAVGLHYFSKNGENSYGNGDMTYARLFGKGTQINSDLTGYYQ 76  
QY 61 WEYFQGNSEGADQATGNKTRLAFAAGLVADYSGDYGRNYGVVYDALGYTDLPEFG 120  
DB 77 WEYFQGNSEGADQATGNKTRLAFAAGLVADYSGDYGRNYGVVYDALGYTDLPEFG 133  
QY 121 DTAYSDFFVGRVGGVATYRNSFFGLVDGLNFAVQYLGKNERDT-----A 166  
DB 134 DTAYSDFFVGRVGGVATYRNSFFGLVDGLNFAVQYLGKNERDT-----A 193  
QY 167 RRSNGDVGGSISYEYEGFGVIGAYGAADRNLQEAQPLGNGKKAQWATGLKYDANNI 225  
DB 194 LRQNGDVGGSISYEYEGFGVIGAYGAADRNLQEAQPLGNGKKAQWATGLKYDANNI 253  
QY 226 YLAANGETRNATPIITNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYKSKADV-E 284  
DB 254 YLAANGETRNATPIITNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYKSKADV-E 308  
QY 285 GIGDVLVNYFEVGATYFYNKNNSTVYDIINOIDSNNKL-----GVGSDTVA 340  
DB 309 GIGDVLVNYFEVGATYFYNKNNSTVYDIINOIDSNNKL-----GVGSDTVA 367

RESULT 4  
US-09-041-889-29  
Sequence 29, Application US/09041889  
Patent No. 6033864  
GENERAL INFORMATION:  
APPLICANT: Braun, Jonathan  
APPLICANT: Cohavy, Offer  
TITLE OF INVENTION: Diagnosis, Prevention and Treatment of  
TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes Thereof, Using  
TITLE OF INVENTION: Microbial UC PANCA antigens  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/041,889  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/837,058  
FILING DATE: 11-APR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P-PM 3006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 377 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-041-889-29

Query Match 56.9%; Score 1029; DB 3; Length 377;  
Best Local Similarity 56.9%; Pred. No. 4.8e-88;  
Matches 209; Conservative 41; Mismatches 79; Indels 38; Gaps 9;  
QY 1 AEYNDGKNDVLYGKAVGLHYFSKNGENSYGNGDMTYARLFGKGTQINSDLTGYYQ 60  
DB 22 AEYNDGKNDVLYGKAVGLHYFSKNGENSYGNGDMTYARLFGKGTQINSDLTGYYQ 76  
QY 61 WEYFQGNSEGADQATGNKTRLAFAAGLVADYSGDYGRNYGVVYDALGYTDLPEFG 120  
DB 77 WEYFQGNSEGADQATGNKTRLAFAAGLVADYSGDYGRNYGVVYDALGYTDLPEFG 134  
QY 121 DTAYSDFFVGRVGGVATYRNSFFGLVDGLNFAVQYLGKNE-----RDTAR 167  
DB 135 DTAYSDFFVGRVGGVATYRNSFFGLVDGLNFAVQYLGKNE-----RDTAR 193  
QY 168 RRSNGDVGGSISYEYEGFGVIGAYGAADRNLQEAQPLGNGKKAQWATGLKYDANNI 226  
DB 194 LRQNGDVGGSISYEYEGFGVIGAYGAADRNLQEAQPLGNGKKAQWATGLKYDANNI 253  
QY 227 LAANGETRNATPIITNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYKSKADV-E 284  
DB 254 LAANGETRNATPIITNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYKSKADV-E 310  
QY 285 GIGDVLVNYFEVGATYFYNKNNSTVYDIINOIDSNNKL-----GVGSDTVA 333  
DB 311 GIGDVLVNYFEVGATYFYNKNNSTVYDIINOIDSNNKL-----GVGSDTVA 370  
QY 334 VGIVYQF 340  
DB 371 LGLVYQF 377

RESULT 5  
US-09-041-889-28  
Sequence 28, Application US/09041889  
Patent No. 6033864  
GENERAL INFORMATION:  
APPLICANT: Braun, Jonathan  
APPLICANT: Cohavy, Offer  
TITLE OF INVENTION: Diagnosis, Prevention and Treatment of  
TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes Thereof, Using  
TITLE OF INVENTION: Microbial UC PANCA antigens  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/041,889  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: US 08/837,058
; FILING DATE: 11-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-M 3006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 323 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-041-889-2R

```

Query Match	47.58;	Score	858.5;	DB	3;	Length	323;
Best Local Similarity	57.28;	Pred.	No. 3.le-72;				
Matches	174;	Conservative	33;	Mismatches	72;	Indels	25;
Gaps	7;						
QY	1	AEYKDGKNDVLGKAVGLHYFSGKNGCNSYGGNGDMTYARLGKGETQINSDLTYGQ	60				
Db	22	AEYKDGKNDLVLGKVDGLHYFSDNSAK-----DGOQSYARLGKGETQINDLTYGQ	76				
QY	61	WEYNFQGNNSGADAQTGNKTRLAFAGLKYADVGSFDYGRNYGVVYDALGYTMDLPEFGG	120				
Db	77	WEYNIQANNTESSKNQSW--TRLAFAGLKFDYGSFDYGRNYGVWYDIEGWTMDLPEFGG	134				
QY	121	DT-AYSDDFFVGRGVGATYRNSNFFGLVDGLNFAVQYLGKNE-----RDRTAR	167				
Db	135	DSVTNADNFMTGRANGVATYRNTDFGLVNLGNLFAVQYQGNNEGASNGQETNNGRD--VR	193				
QY	168	RSNGDVGGSISIEY--EGFGVLGAYGAADRTNLQEAOPLGNGKKAQWATGLKYDANNIY	226				
Db	194	HENGDGWGLSTTYDLGMGFSAGAAATSDSRDINDQVNHTAAGCDRADAWTAGLKYDANNIY	253				
QY	227	LAANYGTRNATPITNKFTNTSGFANKTODVLLVAOYQFDFGLRPSYATYTKSKAKDVEGI	286				
Db	254	LATWYSTRNNTPGD---SDYAVANKTQNEFVATAGYQFDFGLRPAVSFLMSKGRDLHAA	310				
QY	287	GDVD	290				
Db	311	GGAD	314				

```

1  RESULT 6
2  US-08-096-182A-4
3  ; Sequence 4, Application US/08096182A
4  ; Patent No. 5439808
5  ; GENERAL INFORMATION:
6  ; APPLICANT: Blake, Milan S.
7  ; APPLICANT: Tai, Joseph Y.
8  ; APPLICANT: Qi, Huilin L.
9  ; APPLICANT: Liang, Shu-Mei
10 ; APPLICANT: Hronowski, Lucjan J.J.
11 ; APPLICANT: Pullen, Jeffrey K.
12 ; TITLE OF INVENTION: Method for the High Level Expression,
13 ; TITLE OF INVENTION: Purification and Refolding of the Outer Membrane Group B
14 ; TITLE OF INVENTION: Porin Proteins from Neisseria Meningitidis
15 ; NUMBER OF SEQUENCES: 23
16 ; CORRESPONDENCE ADDRESS:
17 ; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
18 ; STREET: 1100 New York Ave., Suite 600
19 ; CITY: Washington
20 ; STATE: D.C.
21 ; COUNTRY: USA
22 ; ZIP: 20005
23 ; COMPUTER READABLE FORM:
24 ; MEDIUM TYPE: Floppy disk
25 ; COMPUTER: IBM PC compatible
26 ; OPERATING SYSTEM: PC-DOS/MS-DOS
27 ; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/096,182A
; FILING DATE: 23-JUL-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438,0060000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 342 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-096-182A-4

Query Match      8.7%; Score 157; DB 1; Length 342;
Best Local Similarity 24.7%; Pred. No. 8.5e-07;
Matches 83; Conservative 45; Mismatches 150; Indels 58; Gaps 16;

QY   11 VDLGKAVGLHYHFSKGNGENSYGNGDMTY-----ARLGFKGTQINSDLTGYGOW 61
     ||||| :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: |
Db   3 VTLYGTIKAGVEYSRVKADCTYKAQGCKSKTATQIADFSGKIGFKGGEDLGNMGKAIWQL 62

QY   62 EYNFGNNSGDAQDTGNKTRLPAGLKAYADVGSFDYGR--NYGVVYDALGYDMLPEFG 119
     ||||| :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: |
Db   63 E---OKASIAITGNSGWN--RQSPILGL----GGFCIVRAGNLNTVLKDSGDNVNAWESG 113

QY   120 GDTAYSDDFFVRGVGGVATYRNSNFFG-LVDGLNEFAVQYL GK--NERDTARRSNGDV 174
     :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: |
Db   114 SNT--EDVLGLTGIRVESREISRYDSPVFAGSGSVQVPVRDNANDVKYHKTKSSRE 171

QY   175 GGSISYEYEGFIGVGAYGA-----ADR--TNLQAQPLGNCKKAEQWATGLKYD 221
     :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: |
Db   172 SYHAGLAYENAFGPGQYAGSFAKYADLTDAERVAVNTANHPV---KDQVHRVVAGID 228

QY   222 ANNIY--LAANYGETRNATPITNKFTWTSFANKTDQVLLVAQYQDFGLRPISIATKSK 279
     ||||| :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: |
Db   229 ANDLYSVAGQYEAKN-----NEVGSTKKKHKEQTQVAATAAYREG-NVTPRVSYAHGF 282

QY   280 AKDVEGGDLDLVNY-----FEVGATYYFNKNMSTYV 311
     :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: |
Db   283 KAKVNGYKD---ANYQDVIVGADYDFSRKTSALV 315


RESULT 7
US-08-877-109-4
: Sequence 4, Application US/08877109
: Patent No. 5747287
: GENERAL INFORMATION:
: APPLICANT: Blake, Milan S.
: APPLICANT: Tai, Joseph Y.
: APPLICANT: Qi, Hulin L.
: APPLICANT: Liang, Shu-Mei
: APPLICANT: Hronowski, Lucjan J.J.
: APPLICANT: Pullen, Jeffrey K.
: TITLE OF INVENTION: Method for the High Level Expression,
: PURATION AND REFOOLDING OF THE OUTER MEMBRANE PROTEIN
: FROM NEISSERIA MENINGITIDIS
: TITLE OF INVENTION: Purification and Refolding of the Outer Membrane Protein
: FROM NEISSERIA MENINGITIDIS
: NUMBER OF SEQUENCES: 23
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sterne, Kessler, Goldstein & Fox
: STREET: 1100 New York Ave., NW
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/877,109  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/431,264  
FILING DATE: 28-APR-1995  
APPLICATION NUMBER: 08/096,182  
FILING DATE: 23-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Esmond, Robert W.  
REGISTRATION NUMBER: 32,893  
REFERENCE/DOCKET NUMBER: 1438.0060001  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 342 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-877-109-4

Query Match 8.7%; Score 157; DB 1; Length 342;  
Best Local Similarity 24.7%; Pred. No. 8.5e-07;  
Matches 83; Conservative 45; Mismatches 150; Indels 58; Gaps 16;  
QY 11 VDLYGRAVLHLYFSKNGNSYSGNGDMTY-----ARLGFKGQTOINSLTGYGQW 61  
DB 3 VTLGTIKAGVEVSRVKDAGTKYKAQGGKSKTATQIADFGSKIGKQEDLGNMKAIWQL 62  
QY 62 EYNFQNNSEGADQGNKTRLAFAGLYADVGSFDYGR--NYGVYDALGYTDMLEPFG 119  
DB 63 E---QKASITAGTNSGNGN--RQSFGLK---GGFCTVRAGNLNTLYKDSGDNVNAWESG 113  
QY 120 GDTAYSDDFVGRVGVVATYRNSNFFG--LVQGLNFAVQYLKG---NERDTARRSNGDGV 174  
DB 114 SNT--EDVLGLGTIGRVSREISRYVDSVPVAFGSGVQVYPRDNANDVDKYKHTKSSRE 171  
QY 175 GGSISYEYEGFIVGAYGA-----ADR--TNLQEAOLPLNGKKAQWATGLKYD 221  
DB 172 SYHAGLYENAGPFGQYAGSFAKYADLNTDAERVAVNTANAHVP---KDYQVHRVVAGYD 228  
QY 222 ANNIY--LAANYGETRNATPITNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATKSK 279  
DB 229 ANDLYSVAGQYEAAKN-----NEVGSTKGKHEQTOVAATAAYRFG-NVTPRVSYAHGF 282  
QY 280 AKDVEGIGDVLVNY-----FEVGATYFFNKNMSTYV 311  
DB 283 KAKVNGVKD---ANYQYDQVIVGADYDFSKRTSALV 315

RESULT 8  
US-08-798-760-4  
Sequence 4, Application US/08/98760  
Patent No. 6013267  
GENERAL INFORMATION:  
APPLICANT: Blake, Milan S.  
APPLICANT: Tai, Joseph Y.  
APPLICANT: Qi, Huilin L.  
APPLICANT: Liang, Shu-Mei  
APPLICANT: Hronowski, Lucjan J.J.  
APPLICANT: Pullen, Jeffrey K.  
TITLE OF INVENTION: Method for the High Level Expression,  
Purification and Refolding of the Outer Membrane Group B  
PORIN PROTEINS FROM NEISSERIA MENINGITIDIS  
TITLE OF INVENTION:  
NUMBER OF INVENTIONS: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1100 New York Ave., Suite 600

CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/798,760  
FILING DATE: 11-FEB-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Esmond, Robert W.  
REGISTRATION NUMBER: 32,893  
REFERENCE/DOCKET NUMBER: 1438.0060002  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 342 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-798-760-4

Query Match 8.7%; Score 157; DB 3; Length 342;  
Best Local Similarity 24.7%; Pred. No. 8.5e-07;  
Matches 83; Conservative 45; Mismatches 150; Indels 58; Gaps 16;  
QY 11 VDLYGRAVLHLYFSKNGNSYSGNGDMTY-----ARLGFKGQTOINSLTGYGQW 61  
DB 3 VTLGTIKAGVEVSRVKDAGTKYKAQGGKSKTATQIADFGSKIGKQEDLGNMKAIWQL 62  
QY 62 EYNFQNNSEGADQGNKTRLAFAGLYADVGSFDYGR--NYGVYDALGYTDMLEPFG 119  
DB 63 E---QKASITAGTNSGNGN--RQSFGLK---GGFCTVRAGNLNTLYKDSGDNVNAWESG 113  
QY 120 GDTAYSDDFVGRVGVVATYRNSNFFG--LVQGLNFAVQYLKG---NERDTARRSNGDGV 174  
DB 114 SNT--EDVLGLGTIGRVSREISRYVDSVPVAFGSGVQVYPRDNANDVDKYKHTKSSRE 171  
QY 175 GGSISYEYEGFIVGAYGA-----ADR--TNLQEAOLPLNGKKAQWATGLKYD 221  
DB 172 SYHAGLYENAGPFGQYAGSFAKYADLNTDAERVAVNTANAHVP---KDYQVHRVVAGYD 228  
QY 222 ANNIY--LAANYGETRNATPITNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATKSK 279  
DB 229 ANDLYSVAGQYEAAKN-----NEVGSTKGKHEQTOVAATAAYRFG-NVTPRVSYAHGF 282  
QY 280 AKDVEGIGDVLVNY-----FEVGATYFFNKNMSTYV 311  
DB 283 KAKVNGVKD---ANYQYDQVIVGADYDFSKRTSALV 315

RESULT 9  
PCT-US94-08327-4  
Sequence 4, Application PC/TUS9408327  
GENERAL INFORMATION:  
APPLICANT: The Rockefeller University  
APPLICANT: 1230 York Avenue  
APPLICANT: New York, New York 10021  
APPLICANT: United States of America  
APPLICANT: 12103 Indian Creek Court  
APPLICANT: Beltsville, Maryland 20705  
APPLICANT: United States of America  
APPLICANT: Tai, Joseph Y.  
APPLICANT: Qi, Huilin L.  
APPLICANT: Liang, Shu-Mei  
APPLICANT: Hronowski, Lucjan J.J.  
APPLICANT: Pullen, Jeffrey K.

;; TITLE OF INVENTION: Method for the High Level  
;; TITLE OF INVENTION: Expression,  
;; TITLE OF INVENTION: Purification and Refolding of the Outer Membrane  
;; TITLE OF INVENTION: Group B  
;; TITLE OF INVENTION: Porin Proteins from Neisseria Meningitidis  
;; NUMBER OF SEQUENCES: 23  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
;; STREET: 1100 New York Ave., Suite 600  
;; CITY: Washington  
;; STATE: D.C.  
;; COUNTRY: USA  
;; ZIP: 20005-3934  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US94/08327  
;; FILING DATE: Herewith  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/096,182  
;; FILING DATE: 23 July 1993  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Esmond, Robert W.  
;; REGISTRATION NUMBER: 32,893  
;; REFERENCE/DOCKET NUMBER: 1438.006PC00  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202) 371-2600  
;; TELEFAX: (202) 371-2540  
;; INFORMATION FOR SEQ ID NO: 4:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 342 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
PCT-US94-08327-4

Query Match 8.7%; Score 157; DB 5; Length 342;  
Best Local Similarity 24.7%; Pred. No. 8.5e-07;  
Matches 83; Conservative 45; Mismatches 150; Indels 58; Gaps 16;  
Qy 11 VDLGKAVGLHYFSGKNGENSYGGNGDMTY-----ARLGFKEGTQINSDLTGYGQW 61  
Db 3 VTLGKAGVEVSRVKDAGTYKAQGGKSKTATQIADFGSKIGFGQEDLGNMKAIWQL 62  
Qy 62 EYNFGNNSGADAGTGNKTRLAFLAGLYADVGSFDYGR--NYGVVYDALGYTDMLEPFG 119  
Db 63 E---QKASIAGTNSGWN--RQSFGLK-----GGFGTVRAGNLNTVLKDSGDNVNAWESG 113  
Qy 120 GDTAYSDFFVGRGVGATYRNSNFFG--LVDGLNFAYQYLK-----NERDTARRSNGDGV 174  
Db 114 SNT--EDVLGLGTIGRVESREISRYDSPVAGFSGVQYVPRNDANDVYKHKTKSRE 171  
Qy 175 GGSISYEYEGFGIVGAYGA-----ADR--TNLQEAQPLGNGKAKQWATGLKYD 221  
Db 172 SYHAGLKYENAGFFGQYAGSFAYKADLNTDAERVAVNTANAHPV---KDYQVHRVYAGYD 228  
Qy 222 ANNIY--LAANYGETRNATPTITNKFTNTSGFANKTQDVLVLAQYQFDGLRPSIATYKSK 279  
Db 229 ANDLVSVAGQYEAANK-----NEVGSTKGKKHEQTQVAATAAAYRFG-NVTPRVSYAHGF 282  
Qy 280 AKDEGIGDVLVNY-----FEVGATYYFNKNMSTVV 311  
Db 283 KAKYNGVKD---ANYQYQVIVGADYDFSKRTSALV 315

RESULT 10  
US-08-096-182A-6  
; Sequence 6, Application US/08096182A  
; Patent No. 5439808

;; GENERAL INFORMATION:  
;; APPLICANT: Blake, Milan S.  
;; APPLICANT: Tai, Joseph Y.  
;; APPLICANT: Qi, Huilin L.  
;; APPLICANT: Liang, Shu-Mei  
;; APPLICANT: Hronowski, Lucjan J.J.  
;; APPLICANT: Pullen, Jeffrey K.  
;; TITLE OF INVENTION: Method for the High Level Expression,  
;; TITLE OF INVENTION: Purification and Refolding of the Outer Membrane Group B  
;; TITLE OF INVENTION: Porin Proteins from Neisseria Meningitidis  
;; NUMBER OF SEQUENCES: 23  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
;; STREET: 1100 New York Ave., Suite 600  
;; CITY: Washington  
;; STATE: D.C.  
;; COUNTRY: USA  
;; ZIP: 20005  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/096,182A  
;; FILING DATE: 23-JUL-1993  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Esmond, Robert W.  
;; REGISTRATION NUMBER: 32,893  
;; REFERENCE/DOCKET NUMBER: 1438.0060000  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202) 371-2600  
;; TELEFAX: (202) 371-2540  
;; INFORMATION FOR SEQ ID NO: 5:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 363 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-096-182A-6

Query Match 8.7%; Score 157; DB 1; Length 363;  
Best Local Similarity 24.7%; Pred. No. 9.2e-07;  
Matches 83; Conservative 45; Mismatches 150; Indels 58; Gaps 16;  
Qy 11 VDLGKAVGLHYFSGKNGENSYGGNGDMTY-----ARLGFKEGTQINSDLTGYGQW 61  
Db 24 VTLGKAGVEVSRVKDAGTYKAQGGKSKTATQIADFGSKIGFGQEDLGNMKAIWQL 83  
Qy 62 EYNFGNNSGADAGTGNKTRLAFLAGLYADVGSFDYGR--NYGVVYDALGYTDMLEPFG 119  
Db 84 E---QKASIAGTNSGWN--RQSFGLK-----GGFGTVRAGNLNTVLKDSGDNVNAWESG 134  
Qy 120 GDTAYSDFFVGRGVGATYRNSNFFG--LVDGLNFAYQYLK-----NERDTARRSNGDGV 174  
Db 135 SNT--EDVLGLGTIGRVESREISRYDSPVAGFSGVQYVPRNDANDVYKHKTKSRE 192  
Qy 175 GGSISYEYEGFGIVGAYGA-----ADR--TNLQEAQPLGNGKAKQWATGLKYD 221  
Db 193 SYHAGLKYENAGFFGQYAGSFAYKADLNTDAERVAVNTANAHPV---KDYQVHRVYAGYD 249  
Qy 222 ANNIY--LAANYGETRNATPTITNKFTNTSGFANKTQDVLVLAQYQFDGLRPSIATYKSK 279  
Db 250 ANDLVSVAGQYEAANK-----NEVGSTKGKKHEQTQVAATAAAYRFG-NVTPRVSYAHGF 303  
Qy 280 AKDEGIGDVLVNY-----FEVGATYYFNKNMSTVV 311  
Db 304 KAKYNGVKD---ANYQYQVIVGADYDFSKRTSALV 336

RESULT 11  
US-08-877-109-6



; Sequence 6, Application US/08877109

; Patent No. 5747287

; GENERAL INFORMATION:

; APPLICANT: Blake, Milan S.

; APPLICANT: Tai, Joseph Y.

; APPLICANT: Qi, Huilin L.

; APPLICANT: Liang, Shu-Mei

; APPLICANT: Hronowski, Lucjan J.J.

; APPLICANT: Pullen, Jeffrey K.

; TITLE OF INVENTION: Method for the High Level Expression,

; TITLE OF INVENTION: Purification and Refolding of the Outer Membrane Group B

; TITLE OF INVENTION: Porin Proteins from Neisseria meningitidis

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox

; STREET: 1100 New York Ave., NW

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: PC-DOS/MS-DOS

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/877,109

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/431,264

; FILING DATE: 28-APR-1995

; APPLICATION NUMBER: 08/096,182

; FILING DATE: 23-JUL-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Esmond, Robert W.

; REGISTRATION NUMBER: 32,893

; REFERENCE/DOCKET NUMBER: 1438.0060001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 371-2600

; TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 363 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-877-109-6

Query Match 8.7%; Score 157; DB 1; Length 363;

Best Local Similarity 24.7%; Pred. No. 9.2e-07;

Matches 83; Conservative 45; Mismatches 150; Indels 58; Gaps 16;

QY 11 VDLGKAVGLHYFSKGENSYGNGDMTY-----ARLFGKGTQNSDLTGYGQW 61

Db 24 VTLGTTIKAGVSRVKDAGTYKAQGGSKTATQIADFGSKIGFKGQEDLGNMGKAIWL 83

QY 62 EYFQGNSSGADAQTKNTRLAFAGLKYADVGSFDYGR--NYGVVYDALGYTDMLPFEG 119

Db 84 E---QKASIACTNSGWN--RQSFGLK-----GGFGTVRAGNLTVLKDSGDVNAWESG 134

QY 120 GDTAYSDDFVGRVGGVATYRNSNFFG--LVDGLNFAVQYLGR---NERDTARRSNGDV 174

Db 135 SNT--EDVLGLGTIGRVERSEISVRVDSVPFAGSGVQVPRDNANDVKYKHTKSSRE 192

QY 175 GGSISYEGFVGAYGA-----ADR--TNLQEAQPLGNKKAEQWATGLKYD 221

Db 193 SYHAGLKYENAGFFGQYAGSFAYADLNTDAERVAVNTANAHPV---KDYQVHRVAGYD 249

QY 222 ANNIY--LAANYGETRNPATITNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYTKSK 279

Db 250 ANDLYSVAGQYEAANK-----NEVGSTKGGKKEQQTQVAATAAYREG--NVTPRVSAHGF 303

QY 280 AKDVEGIGDVLVNY-----FEVGATYYFKNKMSTYV 311

Db 304 KAKVNGVKD---ANYQYDQVIVGADYDFSKRTSALV 336

RESULT 12

US-08-798-760-6

; Sequence 6, Application US/08798760

; Patent No. 6013267

; GENERAL INFORMATION:

; APPLICANT: Blake, Milan S.

; APPLICANT: Tai, Joseph Y.

; APPLICANT: Qi, Huilin L.

; APPLICANT: Liang, Shu-Mei

; APPLICANT: Hronowski, Lucjan J.J.

; APPLICANT: Pullen, Jeffrey K.

; TITLE OF INVENTION: Method for the High Level Expression,

; TITLE OF INVENTION: Purification and Refolding of the Outer Membrane Group B

; TITLE OF INVENTION: Porin Proteins from Neisseria Meningitidis

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox

; STREET: 1100 New York Ave., Suite 600

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: PC-DOS/MS-DOS

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/798,760

; FILING DATE: 11-FEB-1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Esmond, Robert W.

; REGISTRATION NUMBER: 32,893

; REFERENCE/DOCKET NUMBER: 1438.0060002

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 371-2600

; TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 363 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-798-760-6

Query Match 8.7%; Score 157; DB 3; Length 363;

Best Local Similarity 24.7%; Pred. No. 9.2e-07;

Matches 83; Conservative 45; Mismatches 150; Indels 58; Gaps 16;

QY 11 VDLGKAVGLHYFSKGENSYGNGDMTY-----ARLFGKGTQNSDLTGYGQW 61

Db 24 VTLGTTIKAGVSRVKDAGTYKAQGGSKTATQIADFGSKIGFKGQEDLGNMGKAIWL 83

QY 62 EYFQGNSSGADAQTKNTRLAFAGLKYADVGSFDYGR--NYGVVYDALGYTDMLPFEG 119

Db 84 E---QKASIACTNSGWN--RQSFGLK-----GGFGTVRAGNLTVLKDSGDVNAWESG 134

QY 120 GDTAYSDDFVGRVGGVATYRNSNFFG--LVDGLNFAVQYLGR---NERDTARRSNGDV 174

Db 135 SNT--EDVLGLGTIGRVERSEISVRVDSVPFAGSGVQVPRDNANDVKYKHTKSSRE 192

QY 175 GGSISYEGFVGAYGA-----ADR--TNLQEAQPLGNKKAEQWATGLKYD 221

Db 193 SYHAGLKYENAGFFGQYAGSFAYADLNTDAERVAVNTANAHPV---KDYQVHRVAGYD 249

QY 222 ANNIY--LAANYGETRNPATITNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYTKSK 279

Db 250 ANDLYSVAGQYEAANK-----NEVGSTKGKKHEQTQVAAATAAYRFG-NVTPRVSYAHGF 303

QY 280 AKDVEGIGDVLVNY-----FEVGATYYFNKNMSTYV 311

Db 304 KAKVNGVKD---ANYQYDQVIGADYDFSKRTSALV 336

RESULT 13

PCT-US94-08327-6

; Sequence 6, Application PC/TUS9408327

; GENERAL INFORMATION:

; APPLICANT: The Rockefeller University

; APPLICANT: 1230 York Avenue

; APPLICANT: New York, New York 10021

; APPLICANT: United States of America

; APPLICANT: 12103 Indian Creek Court

; APPLICANT: Beltsville, Maryland 20705

; APPLICANT: United States of America

; APPLICANT: Tai, Joseph Y.

; APPLICANT: Qi, Huilin L.

; APPLICANT: Liang, Shu-Mei

; APPLICANT: Hronowski, Lucjan J.J.

; APPLICANT: Pullen, Jeffrey K.

; TITLE OF INVENTION: Expression,

; TITLE OF INVENTION: Method for the High Level

; TITLE OF INVENTION: Purification and Refolding of the Outer Membrane

; TITLE OF INVENTION: Group B

; TITLE OF INVENTION: Porin Proteins from Neisseria Meningitidis

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox

; STREET: 1100 New York Ave., Suite 600

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005-3934

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US94/08327

; FILING DATE: Herewith

; CLASSIFICATION:

; PRIOR APPLICATION NUMBER: US 08/096,182

; FILING DATE: 23 July 1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Esmond, Robert W.

; REGISTRATION NUMBER: 32,893

; REFERENCE/DOCKET NUMBER: 1438.006PC00

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 371-2600

; TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 363 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

PCT-US94-08327-6

Query Match 8.7%; Score 157; DB 5; Length 363;

Best Local Similarity 24.7%; Pred. No. 9.2e-07;

Matches 83; Conservative 45; Mismatches 150; Indels 58; Gaps 16;

QY 11 VDLGKAVCLHYFKNGENSYGGNGDMTY-----ARLGFKGETQINSDLTGYGQW 61

Db 24 VTLGTTIKAGVEVRKVDAGYTKAQGGKSKATQIADFGSKIGFGQEDLGNMKAIWQL 83

QY 62 EYNFGQNSGEGDAQTGNKTKLAFAGLAYADVGSFDYGR--NYGVVYDALGYTDLPEFG 119

Db 84 E---QKASIAGTNSGWN--RQSFTGLK-----GGFGTVRAGNLNTVLKDSGDNVANWESG 134

QY 120 GDTAYSDDFEFGVGVGATYRNSNEFG--LVDGLNFAYQYLGK---NERDTARRSNGDGV 174

Db 135 SNT--EDVLGLGTIGRVESREISVRYDSPVFAGFSGSVQYVPRDNANDVDKVKHKSSRE 192

QY 175 GGSISYEYEGFIVGAYGA-----ADR--TNLQEAQPLNGKKAEQMATGLKYD 221

Db 193 SYHAGLKYENAGFFQYAGSFAKYADLNTDAERVAVNTANAHPV---KDYQVHRVVAGVD 249

QY 222 ANNIY--LAANYGETRNATPTINKETNTSGFANKTQDVLVLAQYQDFGLRPSIATYTSK 279

Db 250 ANDLYSVAGQYEAANK-----NEVGSTKGKKHEQTQVAAATAAYRFG-NVTPRVSYAHGF 303

QY 280 AKDVEGIGDVLVNY-----FEVGATYYFNKNMSTYV 311

Db 304 KAKVNGVKD---ANYQYDQVIGADYDFSKRTSALV 336

RESULT 14

US-08-096-182A-2

; Sequence 2, Application US/08096182A

; Patent No. 5439808

; GENERAL INFORMATION:

; APPLICANT: Blake, Milan S.

; APPLICANT: Tai, Joseph Y.

; APPLICANT: Qi, Huilin L.

; APPLICANT: Liang, Shu-Mei

; APPLICANT: Hronowski, Lucjan J.J.

; APPLICANT: Pullen, Jeffrey K.

; TITLE OF INVENTION: Method for the High Level Expression,

; TITLE OF INVENTION: Purification and Refolding of the Outer Membrane

; TITLE OF INVENTION: Porin Proteins from Neisseria Meningitidis

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox

; STREET: 1100 New York Ave., Suite 600

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/096,182A

; FILING DATE: 23-JUL-1993

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Esmond, Robert W.

; REGISTRATION NUMBER: 32,893

; REFERENCE/DOCKET NUMBER: 1438.0060000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 371-2600

; TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 309 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-096-182A-2

Query Match 8.2%; Score 147.5; DB 1; Length 309;

Best Local Similarity 25.1%; Pred. No. 5.7e-06;

Matches 76; Conservative 46; Mismatches 116; Indels 65; Gaps 17;

QY 35 NGDMTY-----ARLGFKGETQINSDLTGYGOWEYFQGNNSGEGADAQTGNKTRL 83

Db 19 NGQVTEVTATGIVDLGSKIGFKGQEDLGNLKAIWQVE---QKASIAGTSGWGN--RQ 73

QY 84 AFAGLKADYVGSFSDYGRNYGVVYDALGYTDLPEFGD-----TAYSDFFVGRVGG---- 135  
Db 74 SFGLK-GGFGKLRVGRNLNVLKDT-----GDNPDWSKSDYLGYNKIAEPAR 121  
QY 136 --VATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGSGSISYEYEGFIVGAYGA 193  
Db 122 LISVRYDSPFAGLSGSVQYAL-----NDNAGRHNSESYHAGFNKNGGFFV--QYGG 172  
QY 194 ADRTNLQEAQPLGNK-KAQOWATGLKYDANNIYLAANYGETRNPATITNKFTNTSGFAN 252  
Db 173 AYKRHHQVQEGLNIEKYQHRLVSG--YDNDALY-ASVAVQOQDA-----KLTDAASHN 224  
QY 253 KTQDVLVAQYQDFGLRPSIATYKSKAKDVEG-IGDVLVNYFE---VGATYYFNKNMS 308  
Db 225 SQTEVAATLAYREG-NVTPRVSY-----AHGFKGLVDDADIGNEYDVVVGAEDFSKRTS 279  
QY 309 TYV 311  
Db 280 ALV 282

RESULT 15

US-08-877-109-2

; Sequence 2, Application US/08877109

; Patent No. 5747287

; GENERAL INFORMATION:

; APPLICANT: Blake, Milan S.

; APPLICANT: Tai, Joseph Y.

; APPLICANT: Oi, Huilin L.

; APPLICANT: Liang, Shu-Mei

; APPLICANT: Hronowski, Lucjan J.J.

; APPLICANT: Pullen, Jeffrey K.

; TITLE OF INVENTION: Method for the High Level Expression,

; TITLE OF INVENTION: Purification and Refolding of the Outer Membrane Group B

; TITLE OF INVENTION: Porin Proteins from Neisseria meningitidis

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox

; STREET: 1100 New York Ave., NW

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/877,109

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/431,264

; FILING DATE: 28-APR-1995

; APPLICATION NUMBER: 08/096,182

; FILING DATE: 23-JUL-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Esmond, Robert W.

; REGISTRATION NUMBER: 32,893

; REFERENCE/DOCKET NUMBER: 1438.0060001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 371-2600

; TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 309 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-877-109-2

Query Match

8.2%; Score 147.5; DB 1; Length 309;

Best Local Similarity 25.1%; Pred. No. 5.7e-06;  
Matches 76; Conservative 46; Mismatches 116; Indels 65; Gaps 17;  
QY 35 NGDMTY-----ARLFGKETQINSDLTGYGQWEYNFGNNSGADAQTCNKTRL 83  
Db 19 NGQVTEVTTATGIVDLGSKIGFKGQEDLGNGLKAIWQVE---QKASITAGTDSGNGN--RQ 73  
QY 84 AFAGLKADYVGSFSDYGRNYGVVYDALGYTDLPEFGD-----TAYSDFFVGRVGG---- 135  
Db 74 SFGLK-GGFGKLRVGRNLNVLKDT-----GDNPDWSKSDYLGYNKIAEPAR 121  
QY 136 --VATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGSGSISYEYEGFIVGAYGA 193  
Db 122 LISVRYDSPFAGLSGSVQYAL-----NDNAGRHNSESYHAGFNKNGGFFV--QYGG 172  
QY 194 ADRTNLQEAQPLGNK-KAQOWATGLKYDANNIYLAANYGETRNPATITNKFTNTSGFAN 252  
Db 173 AYKRHHQVQEGLNIEKYQHRLVSG--YDNDALY-ASVAVQOQDA-----KLTDAASHN 224  
QY 253 KTQDVLVAQYQDFGLRPSIATYKSKAKDVEG-IGDVLVNYFE---VGATYYFNKNMS 308  
Db 225 SQTEVAATLAYREG-NVTPRVSY-----AHGFKGLVDDADIGNEYDVVVGAEDFSKRTS 279  
QY 309 TYV 311  
Db 280 ALV 282

Search completed: December 18, 2002, 16:23:14  
Job time : 8.85111 secs

***This Page Blank (uspto)***

GenCore version 5.1.1.3  
Copyright (C) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:21:10 ; Search time 4.71067 Seconds  
(without alignments)  
1213.493 Million cell updates/sec

Title: US-09-490-291-9  
Perfect score: 1809  
Sequence: 1 AEIYNKDKNKVLYGKAVGL.....NKLGVSDDTVAVGIVYQFA 341

Scoring table: BLOSUM62  
Gapex 10.0 , Gapext 0.5

Searched: 106657 seqs, 16763532 residues

Total number of hits satisfying chosen parameters: 106657

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						Description	
Result No.	Score	Query Match	Length	DB	ID		
1	925.5	51.2	322	10	US-09-905-176-24	Sequence 24, Appl	
2	113	6.2	467	10	US-09-912-020-381	Sequence 381, Appl	
3	112	6.2	207	9	US-10-029-180-70	Sequence 70, Appl	
4	110	6.1	1426	10	US-09-912-020-340	Sequence 340, Appl	
5	104.5	5.8	708	8	US-08-834-666A-2	Sequence 2, Appl1	
6	104	5.7	848	10	US-09-841-132-192	Sequence 192, Appl	
7	104	5.7	1530	10	US-09-841-132-178	Sequence 178, Appl	
8	103	5.7	691	8	US-08-834-666A-12	Sequence 12, Appl	
9	103	5.7	691	8	US-08-834-666A-22	Sequence 22, Appl	
10	102	5.6	711	8	US-08-834-666A-8	Sequence 8, Appl1	
11	101	5.6	1325	10	US-09-741-669-304	Sequence 304, Appl	
12	100	5.5	1531	12	US-10-007-693-98	Sequence 98, Appl	
13	99	5.5	733	8	US-08-834-666A-4	Sequence 4, Appl1	
14	98.5	5.4	597	10	US-09-793-306-146	Sequence 146, Appl	
15	96.5	5.3	440	10	US-09-813-242-5131	Sequence 5131, Ap	
16	96.5	5.3	943	9	US-09-996-634-131	Sequence 131, Appl	
17	95.5	5.3	925	9	US-09-924-097-14	Sequence 14, Appl	
18	95	5.3	745	8	US-08-834-666A-6	Sequence 6, Appl1	
19	93.5	5.2	909	10	US-09-332-226-4	Sequence 4, Appl1	

20	93	5.1	672	8	US-08-834-666A-10	Sequence 10, Appl
21	92.5	5.1	595	10	US-09-881-752A-166	Sequence 166, App
22	92.5	5.1	1536	9	US-10-092-880-2	Sequence 2, Appl1
23	92	5.1	1377	10	US-09-815-242-10384	Sequence 10384, A
24	91.5	5.1	700	10	US-09-881-752A-180	Sequence 180, App
25	90.5	5.0	1430	10	US-09-740-274-6	Sequence 6, Appl1
26	90	5.0	272	10	US-09-119-900-15	Sequence 15, Appl
27	90	5.0	934	8	US-08-837-459-19	Sequence 19, Appl
28	90	5.0	1290	10	US-09-881-752A-138	Sequence 138, App
29	89.5	4.9	815	10	US-09-815-242-5106	Sequence 5106, Ap
30	89	4.9	410	10	US-09-905-176-26	Sequence 26, Appl
31	89	4.9	715	9	US-09-870-759-86	Sequence 86, Appl
32	89	4.9	878	10	US-09-912-020-364	Sequence 364, App
33	89	4.9	2122	9	US-09-813-214A-9	Sequence 9, Appl1
34	88.5	4.9	470	10	US-09-881-752A-174	Sequence 174, App
35	88	4.9	1152	10	US-09-945-265-4	Sequence 4, Appl1
36	88	4.9	1153	10	US-09-350-259-3	Sequence 3, Appl1
37	88	4.9	1974	9	US-09-895-913A-12	Sequence 12, Appl
38	87.5	4.8	839	9	US-10-023-437-23	Sequence 23, Appl
39	87	4.8	271	10	US-09-119-900-8	Sequence 8, Appl1
40	87	4.8	518	9	US-09-976-297-2	Sequence 2, Appl1
41	87	4.8	793	10	US-09-881-752A-362	Sequence 362, App
42	87	4.8	2353	10	US-09-797-862-33	Sequence 33, Appl
43	86	4.8	1600	9	US-10-092-880-10	Sequence 10, Appl
44	85.5	4.7	285	9	US-09-895-913A-138	Sequence 138, App
45	85	4.7	565	9	US-10-108-605-299	Sequence 299, App

ALIGNMENTS

RESULT 1

US-09-905-176-24

; Sequence 24, Application US/09905176

; Patent No. US20020150906A1

; GENERAL INFORMATION:

; APPLICANT: CALIFORNIA INSTITUTE OF TECHNOLOGY

; APPLICANT: Debe, Derek A.

; TITLE OF INVENTION: METHOD FOR DETERMINING THREE-DIMENSIONAL PROTEIN STRUCTURE PRO

; FILE REFERENCE: 265/297

; CURRENT APPLICATION NUMBER: US/09/905,176

; CURRENT FILING DATE: 2002-04-05

; PRIOR APPLICATION NUMBER: US 60/218,016

; PRIOR FILING DATE: 2000-07-12

; NUMBER OF SEQ ID NOS: 26

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 24

; LENGTH: 322

; TYPE: PRT

; ORGANISM: Klebsiella pneumoniae

US-09-905-176-24

Query Match		51.2%	Score 925.5	DB 10	Length 322
Best Local Similarity		57.3%	Pred. No. 1.1e-76		
Matches 192		Conservative 35	Mismatches 81	Indels 27	Gaps 8
QY	10	KVDLYGKAVGLHYFSGKNGENSYCGNGDMTYARLGFKEGTQINSDLTGYGQWYVFGQNN	69		
Db	1	KLDLYGKIDGLHYFS-----DDKXVDGDQTYMRGLGVKGETQINDQLTGYGQWYVQANN	55		
QY	70	SEGADQATQNKTRLAFLAGLYADVGSFQYGRNYGVVYDALGYDMLPEFGDGTAYSDDF	129		
Db	56	TESSSDQAW--TRLAFLAGLFGDAGSFQYGRNYGVVYDVTSWTDVLPFGDGTGYSDNFL	113		
QY	130	VGRVGVVATYRNSEFGLVDGLNFAVQYLGKN-----ERDT-----ARRNSGQVGGIS	179		
Db	114	QSRANGVATYRNDFDFFGLV-GLNFALOYQKNGSVSGEGATNNGRALKNGDGFQSVT	172		
QY	180	YE-YEGFVIGVAYCAADRTNLOEAQPLNGKKAQWATGLKYDANNIYLAANTYLAANTY	238		
Db	173	YDIFDGISAGFAYANSKRDTDQNLGLGEGDHAETYGGLKYDANNIYLATQYTOTYAT	232		



Query Match 5.7%; Score 104; DB 10; Length 848;  
Best Local Similarity 22.6%; Pred. No. 0.11;  
Matches 72; Conservative 39; Mismatches 103; Indels 104; Gaps 20;

QY 29 ENSYGGNDWTYARLGPKEGTQINSIDTGY-----GQWBYNEQGNNSEGADA 75  
:::| | : : : : | : : : |  
| : : : |





Db 651 QHGIELGFKIPTINTNTYFPMGAKLEYRMYSLFLNV 688

RESULT 9

US-08-834-666A-22

; Sequence 22, Application US/08834666A

; Patent No. US20020044949A1

GENERAL INFORMATION:

APPLICANT: Kleanthous, Harold

APPLICANT: Lissolo, Ling

APPLICANT: Tomb, Jean-Francois

APPLICANT: Miller, Charles

APPLICANT: Al-Garawi, Amal

TITLE OF INVENTION: 76 kDa Helicobacter Polypeptides and

TITLE OF INVENTION: Corresponding Polynucleotide Molecules

NUMBER OF SEQUENCES: 64

CORRESPONDENCE ADDRESS:

ADDRESSEE: Clark & Elbing LLP

STREET: 176 Federal Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02110

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

APPLICATION NUMBER: US/08/834,666A

FILING DATE: 01-APR-1997

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Clark, Paul T.

REGISTRATION NUMBER: 30,175

REFERENCE/DOCKET NUMBER: 06132/038001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-428-0200

TELEFAX: 617-428-7045

TELEX:

INFORMATION FOR SEQ ID NO: 22:

SEQUENCE CHARACTERISTICS:

LENGTH: 691 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal

FEATURE:

NAME/KEY: Signal Sequence

LOCATION: 1...18

OTHER INFORMATION:

US-08-834-666A-22

Query Match 5.7%; Score 103; DB 8; Length 691;

Best Local Similarity 21.6%; Pred. No. 0.1;

Matches 60; Conservative 27; Mismatches 69; Indels 122; Gaps 15;

Qy 64 NFGQNN--EGADAQTGNKTRLAPAGLYADVGSFDRYGVVYDGLGYDMLPFGGD 121

Db 506 DFQNNMGAMGIGVQGVK---QFFGKK-----RWGLRY--YGFYD-----N 544

Qy 122 TAYSDDFVGRVGVATY-----RNSNFGVLDGLNFVAVOYLGKNERDTRR 168

Db 545 HAYIKSNFFNSASDVWTVYGVMDALYNFINDKNTNF-----LCKNNKLSV-- 589

Qy 169 SNGDVGSGSISYEYEGGIYGAADRTNLQEAQPLGNGKKAQWATGLKYD---ANNI 225

Db 590 ----GLFG-----GFALACT-----SWLSQQVNLTMMGI 616

Qy 226 YLAANYGETRNPATITNKFTNSGFANKTQDVLVLAQYQDFGLRPSIATYTKSKADVEG 285

Db 617 Y-----NANVSTSNF-----QFLFDLGLRMLNLRPKKSDSHAA 650

Qy 286 IGDVDL-----VNYFE-VGATYYFNKNMSTYVDYI 314

Db 651 QHGIELGFKIPTINTNTYFPMGAKLEYRMYSLFLNV 688

RESULT 10

US-08-834-666A-8

; Sequence 8, Application US/08834666A

; Patent No. US20020044949A1

GENERAL INFORMATION:

APPLICANT: Kleanthous, Harold

APPLICANT: Lissolo, Ling

APPLICANT: Tomb, Jean-Francois

APPLICANT: Miller, Charles

APPLICANT: Al-Garawi, Amal

TITLE OF INVENTION: 76 kDa Helicobacter Polypeptides and

TITLE OF INVENTION: Corresponding Polynucleotide Molecules

NUMBER OF SEQUENCES: 64

CORRESPONDENCE ADDRESS:

ADDRESSEE: Clark & Elbing LLP

STREET: 176 Federal Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02110

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

APPLICATION NUMBER: US/08/834,666A

FILING DATE: 01-APR-1997

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Clark, Paul T.

REGISTRATION NUMBER: 30,175

REFERENCE/DOCKET NUMBER: 06132/038001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-428-0200

TELEFAX: 617-428-7045

TELEX:

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 711 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal

FEATURE:

NAME/KEY: Signal Sequence

LOCATION: 1..20

OTHER INFORMATION:

US-08-834-666A-8

Query Match 5.6%; Score 102; DB 8; Length 711;

Best Local Similarity 20.8%; Pred. No. 0.13;

Matches 46; Conservative 26; Mismatches 75; Indels 74; Gaps 8;

Qy 132 RVGVAIYRNSNFTGLVDGLNFAV---QYLGKNERDTRRNSGVDGSGSISYEYEGFIV 188

Db 524 RAGLIAASTTNN--GAMNGIGFQVGYQKQFFGKKNRWGAR----- 560

Qy 189 GAYCAADRTNLQEAQPLGNGKKAQWATGLKYD-----A 222

Db 561 -YGFVDINHTYFNKSNQFFN-SDSDVWTVYGVGSDLLVNFINDKATKHNKISFGAFGGIQLA 618

Qy 223 NNIYLAANYGETRNPATITNKFTNSGFANKTQDVLVLAQYQDFGLRPSIATYTKSKAD 282

Db 619 GTSWLSQVYNLANVNNYKAKINTSNF-----QFLNLGLRLNLRNKRIGAD 667



STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
FEATURE:  
NAME/KEY: Signal Sequence  
LOCATION: 1...20  
OTHER INFORMATION:  
US-08-834-666A-4

Query Match 5.5%; Score 99; DB 8; Length 733;  
Best Local Similarity 20.4%; Pred. No. 0.26;  
Matches 66; Conservative 44; Mismatches 121; Indels 92; Gaps 17;

QY 49 TQINSLDTGYGOWEYNFOGNNSEG-ADAQCNKTRLAFAGLKYADVGSFDFGRNRYGVVYD 107  
Db 443 TNLGNSTAHFTGTBQIQ-QAENIADTLNFKSR-----YSELGN-----TYSITT 488  
QY 108 ALGYTMDLPEFGG-----DTAYSDDFVGRVGGVAT--YRNSNFFGLVDGLNFAVQVLGK 160  
Db 489 ALSKVPNAQSIQNVVSKNNPSPQ-----GIETNYLQNQSYNQIQITIN---QELGR 538  
QY 161 NERD-----TARRNG--DGVGGSISYE-----YEGF----- 185  
Db 539 NPERKVGIVNSQTNNGAMNGIGIQVYKQFFGQKRWGARYGFFDYNHAFIKSSFNSA 598  
QY 186 GIVGAYG-AADR-----TNLQEAQPLGNGKKAQOWATGLKYDANNIYLAANYGETRNATPI 240  
Db 599 SDWYTFGFGADALYFNINDKATNFGLNKNNLSGLFGGIALAGTSLNSEYVNLATVNVV 658  
QY 241 TNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIAYTKSKARDVEGIGVDL-----V 292  
Db 659 YNAKMNVANF-----QFLFNGVGRMNLARSKKSDHAAQHGIELGLKIPIINT 707  
QY 293 NYPE-VGATYFNKNMSTYVDYI 314  
Db 708 NYTFMGAEKLYRRLYSVYLNIV 730

RESULT 14  
US-09-793-306-146  
Sequence 146, Application US/09793306  
Patent No. US20020098200A1  
GENERAL INFORMATION:  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Skeiky, Yasir  
APPLICANT: Ovendale, Pamela  
APPLICANT: Jen, Shyian  
APPLICANT: Lodes, Michael  
APPLICANT: Corixa Corporation  
TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy  
FILE REFERENCE: 014058-00874005  
CURRENT APPLICATION NUMBER: US/09793,306  
CURRENT FILING DATE: 2001-02-26  
PRIOR APPLICATION NUMBER: US 60/185,037  
PRIOR FILING DATE: 2000-02-25  
PRIOR APPLICATION NUMBER: US 60/223,828  
PRIOR FILING DATE: 2000-08-08  
NUMBER OF SEQ ID NOS: 164  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 146  
LENGTH: 597  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: mTTC#3-His  
US-09-793-306-146

Query Match 5.4%; Score 98.5; DB 10; Length 597;  
Best Local Similarity 21.6%; Pred. No. 0.22;  
Matches 59; Conservative 30; Mismatches 93; Indels 91; Gaps 14;

QY 26 NGENSYGGNGDMTYARLGFKGETQINSDLT-----GYGOWEYNFOGNNSEGA 73  
Db 310 GSSNIGFNGNNDNLG-LGNNGNNNIGGLTGDNLVGLTALNSGLNLFNGSGNNGNNIGF 368  
QY 74 DAOTCNKTRLAFAGLKYADVGSFDFGRNRYGVVYDALGYTMDLMLPEFGGDTAYSDDFVGRV 133  
Db 369 -FNSGNN-----NVGFNSGNN-----NEGFNGAGDINTGFGNA 401  
QY 134 GCVAT-YRNSNFF-----GLVDGLNFAVQYLGKNERDTARRNSNGDVGCGSISY 180  
Db 402 GDTWTGFGNAGFFNMGIGNAGNEDMGVNGGGSFNV-----GVGNAGNSGVGFGNAGTLNV 456  
QY 181 EYEGFGIV-----CAYGAADRNLQEAQPLGNGKKAQOWA-----TGLK-- 219  
Db 457 GFANAGSINTGFANSGSINTGFGSDGRNT-----CFFGSSVDQSVSSSGFGNTGMSS 509  
QY 220 --YDANNIYLAANYGETRNATPITNKFTNTSGF 250  
Db 510 GFFNTGNV--SAGYGNNGDVGSGINN-TNSGGF 539

RESULT 15  
US-09-815-242-5131  
Sequence 5131, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlson, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5131  
LENGTH: 440  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-815-242-5131

Query Match 5.3%; Score 96.5; DB 10; Length 440;  
Best Local Similarity 21.0%; Pred. No. 0.23;  
Matches 87; Conservative 45; Mismatches 145; Indels 137; Gaps 19;

QY 39 TYARLGFKGETQINSDLTGYGOWEYNFOGNNSEGAQTKNTRLAFAGLKYADVGV---S 95  
Db 52 TDKFESEKLGRLQAD---YGRFD-GYTNNGNTADAAYFRAYLEFGGTAYRDWKYQIN 107  
QY 96 FDYGRNRYGVVYDALGYTD-----MLPEFGGDTAYSDDF----- 129  
Db 108 YLSRNVG--NDSAGYFDEASVYTGFPVNLKFGRFYDFGLEKATSSKQWVTALERLNT 165

```
QY 130 -----VGR-----VGVV-----TYRNSNFFGLVD----- 149
Db 166 YDIADRVNDNVGTGIAOSSVVGMAFLSGVSENNNDTDGDSVKRYNLRGVFAPLHEPG 225
QY 150 -----CLNEFAVOYLGNERNDTARR-----SNGDG--VGGSTSYE--- 181
Db 226 NVVHLGLQYAYRDLEDSAYDTRIRPRMGMRGVSTNGGNDAGSNGNRGLFGGSSAVEGLWK 285
QY 182 -----YEGFIVGAYGAADRTNLQEAQPLGNGKKAQWATGLKYDANNIYLAANYGETR 235
Db 286 DDSVMGLEGAWALGAFSA-----QAEYLRRTVKAERDREDLK--ASGYAQLAY--TL 334
QY 236 NATPITNKFTNTSGFANKTQDVLVVAQYQDFGLRPSIAIYTKSKAKD-----VEGIG 287
Db 335 TGEPRLYKLDGAKFDTIKPENKEIGAWELF-----YRYDSIKVEDDNIIVVDSATREVG 387
QY 288 DVDLVNVEVVGATYYFNKNMSTVVDYIINOIDS-DNKLGVGSDDTVAVGIVYQF 340
Db 388 DAKGKTH-TLCVNWYANEAVKVSANYVAKAKTKISNANGDDSGDGLVMRLQYVF 440
```

Search completed: December 18, 2002, 16:39:31  
Job time : 9.71067 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:13:49 ; Search time 10.0756 Seconds  
(without alignments)  
3253.588 Million cell updates/sec

Title: US-09-490-291-9  
Perfect score: 1809  
Sequence: 1 AEIYNKDGKVKLYCKAVGL.....NKLGVSDDTVAVGIVYQFA 341

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1805	99.8	362	1 MMSCF	outer membrane por
2	1789	98.9	362	2 D90755	outer membrane por
3	1789	98.9	362	2 D85619	outer membrane por
4	1113	61.5	351	2 D90662	outer membrane por
5	1113	61.5	351	2 B85513	outer membrane por
6	1110	61.4	351	1 MMSCE	outer membrane por
7	1082	59.8	371	2 AH0147	outer membrane por
8	1070	59.1	351	2 S25520	outer membrane por
9	1058	58.5	350	2 S25525	outer membrane por
10	1058	58.5	350	2 AF0543	outer membrane por
11	1051	58.1	365	1 MMBPP2	outer membrane por
12	1051	58.1	367	1 MMECPG	outer membrane por
13	1044	57.7	350	2 S36606	outer membrane por
14	1042	57.6	365	1 MMECNC	outer membrane por
15	1035	57.2	377	2 D64888	outer membrane por
16	1033	57.1	367	2 H91016	outer membrane por
17	1033	57.1	367	2 B85861	outer membrane por
18	1031.5	57.0	363	2 AE0616	outer membrane pro
19	1030	56.9	363	2 S51104	outer membrane por
20	1029.5	56.9	363	2 S43159	outer membrane por
21	1027	56.8	383	2 AE0690	outer membrane por
22	1023.5	56.6	378	1 MMEBPC	outer membrane por
23	1023.5	56.6	378	2 AE0789	outer membrane por
24	1000.5	55.3	378	1 A59139	outer membrane por
25	998.5	55.2	366	2 D90888	putative outer memb
26	998.5	55.2	366	2 D85729	probable outer mem
27	992	54.8	360	2 AE0172	outer membrane por
28	976.5	54.0	394	2 S34263	outer membrane por
29	976.5	54.0	394	2 AC0755	outer membrane por

30	960	53.1	374	2 AH0149	outer membrane pro
31	885	48.9	369	2 S70847	outer membrane por
32	770	42.6	315	2 T46993	hypothetical prote
33	649.5	35.9	255	2 F90878	hypothetical prote
34	635.5	35.1	255	2 D85740	partial probable o
35	573.5	31.7	382	2 G84971	ompF-like porin [1
36	539.5	29.8	214	2 F90966	probable outer mem
37	539.5	29.8	224	2 F85814	probable outer mem
38	486	26.9	171	2 H64960	outer membrane por
39	423	23.4	191	2 G90966	hypothetical prote
40	415	22.9	191	2 G85814	probable outer mem
41	372.5	20.6	123	2 G90878	hypothetical prote
42	372.5	20.6	123	2 C85740	partial probable o
43	351	19.4	134	2 B64961	outer membrane por
44	236.5	13.1	350	2 A82299	outer membrane pro
45	233.5	12.9	96	2 C64900	outer membrane por

ALIGNMENTS

RESULT 1

MMECF

outer membrane porin ompF precursor - Escherichia coli (strain K-12)  
N:Alternate names: outer membrane protein 1a; outer membrane protein b  
C:Species: Escherichia coli  
C>Date: 13-Jun-1983 #sequence\_revision 13-Jun-1983 #text\_change 01-Mar-2002  
A:Accession: A93449; A91301; A90314; A25029; H64832; A03431  
R:Inokuchi, K.; Mutoh, N.; Matsuyama, S.; Mizushima, S.  
Nucleic Acids Res. 10, 6957-6968, 1982  
A:Title: Primary structure of the ompF gene that codes for a major outer membrane pro  
A:Reference number: A93449; MUID:83090452; PMID:6294623  
A:Accession: A93449  
A:Molecule type: DNA  
A:Residues: 1-362 <INO>  
A:Cross-references: GB:J01655; GB:M10311; GB:M10312; MID:g147009; PIDN:AAA24244.1; PI  
A:Experimental source: strain K12  
R:Mutoh, N.; Inokuchi, K.; Mizushima, S.  
FEBS Lett. 137, 171-174, 1982  
A:Title: Amino acid sequence of the signal peptide of OmpF, a major outer membrane pr  
A:Reference number: A91301; MUID:82139379; PMID:7037455  
A:Accession: A91301  
A:Molecule type: DNA  
A:Residues: 1-37 <MUT>  
R:Chen, R.; Kramer, C.; Schmidmayr, W.; Chen-Schmeisser, U.; Henning, U.  
Biochem. J. 203, 33-43, 1982  
A:Title: Primary structure of major outer-membrane protein I (ompF protein, porin) of  
A:Reference number: A90314; MUID:82256494; PMID:7049161  
A:Accession: A90314  
A:Molecule type: protein  
A:Residues: 23-87, 'E', 89-138, 'G', 140-283, 'L', 285-362 <CHE>  
R:Nogami, T.; Mizuno, T.; Mizushima, S.  
J. Bacteriol. 164, 797-801, 1985  
A:Title: Construction of a series of ompF-ompC chimeric genes by in vivo homologous r  
A:Reference number: A91809; MUID:86033642; PMID:2997131  
A:Accession: A25029  
A:Molecule type: DNA  
A:Residues: 33-63 <NOG>  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: H64832  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-362 <BLAT>  
A:Cross-references: GB:AE0000195; GB:U00096; MID:g1787156; PIDN:AAC74015.1; PID:g17871  
A:Experimental source: strain K-12, substrain MGL1655  
C:Genetics:  
A:Gene: ompF; toIF; cmIB; coa; cry  
A:Map position: 21 min  
C:Complex: homotrimer

C:Function: <POR>  
A:Description: Forms aqueous channels that facilitate diffusion of small hydrophilic mole  
C:Function: <REC>  
A:Description: receptor for bacteriophage T2  
C:Superfamily: outer membrane protein phoE  
C:Keywords: membrane protein; porin; trimer  
F:1-72/Domain: signal sequence #status predicted <SIG>  
F:723-362/Product: outer membrane porin ompF #status experimental <MAT>  
  
Query Match 99.8%; Score 1805; DB 1; Length 362;  
Best Local Similarity 100.0%; Pred. No. 9.3e-123; Mismatches 0; Indels 0; Gaps 0;  
Matches 340; Conservative 0;  
  
QY 1 AEIYNKDGKVDLYKAVGLHYFSKGENSYGGNGDMTYARLGFKEGTQINSLDTGYGQ 60  
|||||  
DB 23 AEIYNKDGKVDLYKAVGLHYFSKGENSYGGNGDMTYARLGFKEGTQINSLDTGYGQ 82  
|||||  
  
QY 61 WEYNFQGNNSGADATQGNKTRLAFAGLYADVGSFDYGRNYYGVYDMLPFGG 120  
|||||  
DB 83 WEYNFQGNNSGADATQGNKTRLAFAGLYADVGSFDYGRNYYGVYDMLPFGG 142  
|||||  
  
QY 121 DTAYSDDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGSGSISY 180  
|||||  
DB 143 DTAYSDDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGSGSISY 202  
|||||  
  
QY 181 EYEGFIVGAYGAADRTNLQEAQPLNGKKAQOWATGLKYDANNIYLAANYGETRNPPI 240  
|||||  
DB 203 EYEGFIVGAYGAADRTNLQEAQPLNGKKAQOWATGLKYDANNIYLAANYGETRNPPI 262  
|||||  
  
QY 241 TNKFTNTSGFANKTQDVLVAQYQDFGLRPSIATYKSKADVEGIDVDLVNYPEVGAT 300  
|||||  
DB 263 TNKFTNTSGFANKTQDVLVAQYQDFGLRPSIATYKSKADVEGIDVDLVNYPEVGAT 322  
|||||  
  
QY 301 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 340  
|||||  
DB 323 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 362  
|||||  
  
RESULT 2  
D90755  
outer membrane protein la ECs1012 [imported] - Escherichia coli (strain O157:H7, substra  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
C:Accession: D90755  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gend  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: D90755  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-362 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA834435.1; PID:q13360471; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Genetics:  
C:Superfamily: outer membrane protein phoE  
  
Query Match 98.9%; Score 1789; DB 2; Length 362;  
Best Local Similarity 99.4%; Pred. No. 1.3e-121; Mismatches 2; Indels 0; Gaps 0;  
Matches 338; Conservative 0;  
  
QY 1 AEIYNKDGKVDLYKAVGLHYFSKGENSYGGNGDMTYARLGFKEGTQINSLDTGYGQ 60  
|||||  
DB 23 AEIYNKDGKVDLYKAVGLHYFSKGENSYGGNGDMTYARLGFKEGTQINSLDTGYGQ 82  
|||||  
  
QY 61 WEYNFQGNNSGADATQGNKTRLAFAGLYADVGSFDYGRNYYGVYDMLPFGG 120  
|||||  
DB 83 WEYNFQGNNSGADATQGNKTRLAFAGLYADVGSFDYGRNYYGVYDMLPFGG 142  
|||||  
  
QY 121 DTAYSDDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGSGSISY 180  
|||||  
DB 143 DTAYSDDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGSGSISY 202  
|||||  
  
QY 181 EYEGFIVGAYGAADRTNLQEAQPLNGKKAQOWATGLKYDANNIYLAANYGETRNPPI 240  
|||||  
DB 203 EYEGFIVGAYGAADRTNLQEAQPLNGKKAQOWATGLKYDANNIYLAANYGETRNPPI 262  
|||||  
  
QY 241 TNKFTNTSGFANKTQDVLVAQYQDFGLRPSIATYKSKADVEGIDVDLVNYPEVGAT 300  
|||||  
DB 263 TNKFTNTSGFANKTQDVLVAQYQDFGLRPSIATYKSKADVEGIDVDLVNYPEVGAT 322  
|||||  
  
QY 301 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 340  
|||||  
DB 323 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 362  
|||||  
  
RESULT 2  
D90755  
outer membrane protein la ECs1012 [imported] - Escherichia coli (strain O157:H7, substra  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
C:Accession: D90755  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gend  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: D90755  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-362 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA834435.1; PID:q13360471; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Genetics:  
C:Superfamily: outer membrane protein phoE  
  
Query Match 98.9%; Score 1789; DB 2; Length 362;  
Best Local Similarity 99.4%; Pred. No. 1.3e-121; Mismatches 2; Indels 0; Gaps 0;  
Matches 338; Conservative 0;  
  
QY 1 AEIYNKDGKVDLYKAVGLHYFSKGENSYGGNGDMTYARLGFKEGTQINSLDTGYGQ 60  
|||||  
DB 23 AEIYNKDGKVDLYKAVGLHYFSKGENSYGGNGDMTYARLGFKEGTQINSLDTGYGQ 82  
|||||  
  
QY 61 WEYNFQGNNSGADATQGNKTRLAFAGLYADVGSFDYGRNYYGVYDMLPFGG 120  
|||||  
DB 83 WEYNFQGNNSGADATQGNKTRLAFAGLYADVGSFDYGRNYYGVYDMLPFGG 142  
|||||  
  
QY 121 DTAYSDDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGSGSISY 180  
|||||

DB 143 DTAYSDDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGSGSISY 202  
QY 181 EYEGFIVGAYGAADRTNLQEAQPLNGKKAQOWATGLKYDANNIYLAANYGETRNPPI 240  
|||||  
DB 203 EYEGFIVGAYGAADRTNLQEAQPLNGKKAQOWATGLKYDANNIYLAANYGETRNPPI 262  
|||||  
QY 241 TNKFTNTSGFANKTQDVLVAQYQDFGLRPSIATYKSKADVEGIDVDLVNYPEVGAT 300  
|||||  
DB 263 TNKFTNTSGFANKTQDVLVAQYQDFGLRPSIATYKSKADVEGIDVDLVNYPEVGAT 322  
|||||  
  
QY 301 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 340  
|||||  
DB 323 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 362  
|||||  
  
RESULT 3  
B85619  
outer membrane protein la (Ia,b,F) [imported] - Escherichia coli (strain O157:H7, sub  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: B85619  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May  
illet, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: B85619  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-362 <STO>  
A:Cross-references: GB:AE005174; NID:g12514104; PIDN:AA655414.1; GSPDB:GN00145; UMG:  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: ompF  
C:Superfamily: outer membrane protein phoE  
  
Query Match 98.9%; Score 1789; DB 2; Length 362;  
Best Local Similarity 99.4%; Pred. No. 1.3e-121; Mismatches 2; Indels 0; Gaps 0;  
Matches 338; Conservative 0;  
  
QY 1 AEIYNKDGKVDLYKAVGLHYFSKGENSYGGNGDMTYARLGFKEGTQINSLDTGYGQ 60  
|||||  
DB 23 AEIYNKDGKVDLYKAVGLHYFSKGENSYGGNGDMTYARLGFKEGTQINSLDTGYGQ 82  
|||||  
  
QY 61 WEYNFQGNNSGADATQGNKTRLAFAGLYADVGSFDYGRNYYGVYDMLPFGG 120  
|||||  
DB 83 WEYNFQGNNSGADATQGNKTRLAFAGLYADVGSFDYGRNYYGVYDMLPFGG 142  
|||||  
  
QY 121 DTAYSDDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGSGSISY 180  
|||||  
DB 143 DTAYSDDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGSGSISY 202  
|||||  
  
QY 181 EYEGFIVGAYGAADRTNLQEAQPLNGKKAQOWATGLKYDANNIYLAANYGETRNPPI 240  
|||||  
DB 203 EYEGFIVGAYGAADRTNLQEAQPLNGKKAQOWATGLKYDANNIYLAANYGETRNPPI 262  
|||||  
  
QY 241 TNKFTNTSGFANKTQDVLVAQYQDFGLRPSIATYKSKADVEGIDVDLVNYPEVGAT 300  
|||||  
DB 263 TNKFTNTSGFANKTQDVLVAQYQDFGLRPSIATYKSKADVEGIDVDLVNYPEVGAT 322  
|||||  
  
QY 301 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 340  
|||||  
DB 323 YFENKMSYVDYIINQIDSDNKLGVGSDDTVAVGIVYQF 362  
|||||  
  
RESULT 4  
D90662  
outer membrane pore protein PhoE ECs0268 [imported] - Escherichia coli (strain O157:H  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
C:Accession: D90662  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

Db 77 WEAEFAGNKAESDTAQ--QKTRLAFAGLKKYKDLGSGFDYGRNLGALYDVEAWTDMFPEGG 134

QY 121 D-TAYSDDFFVGRVGGVATYRNSNFFGLVDGLNFAVOYLGNKNERDRTARRSNGDGVGGGIS 179

Db 135 DSSAQTDNFWTKRASGLATYRTDFFGVIDGLNLTLYQCKNERNRDYKKQNGDGFGLSLT 194

QY 180 YEYEG--FGIVGAYGAADRNLQBAQPLNGKKAEQWATGLKYDANNIYLAANYGETRPA 237

Db 195 YDFGSDFAISGAYTNSDRTEQNLQSRGTGKRAEAWATGLKYDANNIYLATFYSETRKM 254

QY 238 TPITNKFTNTSGFANKTDQVLLVNAQYQDFDGLRPSIATYKSKAKDVEGIDGVOLVNYFEV 297

Db 255 TPI-----TGGFANKTQNFEAQYQDFDGLRPSLGSYVLSSKGDIEGIDEDLVNIDV 308

QY 298 GATYFYFNKNKSTVYDYIINQIDTSNKLGVGSDDTVAAGVYQF 340

Db 309 GATYFYFNKNKSAFYVDYIKINQLNDSNKLINNDIDIVAGVTYQF 351

RESULT 6

MMECPE

C:Species: Escherichia coli

C:Date: 13-Jun-1983 #sequence\_revision 13-Jun-1983 #text\_change 01-Mar-2002

C:Accession: A03432; B64749

R:Overbeeke, N.; Bergmans, H.; van Mansfeld, F.; Lugtenberg, B.

J. Mol. Biol. 163, 513-532, 1983

A:Title: Complete nucleotide sequence of phoE, the structural gene for the phosphate

A:Reference number: A92893; MUID:83189086; PMID:6341601

A:Accession: A03432

A:Molecule type: DNA

A:Residues: 1-351 <OVE>

A:Cross-references: GB:D83536; NID:g4902908; PIDN:BAW7910.1; PID:g4902976

A:Experimental source: strain K12

A:Note: this is one of the proteins induced when the E. coli cells are grown under phoE

d compounds, and some other negatively charged solutes

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: B64749

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-351 <BLAT>

A:Cross-references: GB:AE000132; GB:U00096; NID:g2367098; PIDN:NAAC73345.1; PID:g17864

A:Experimental source: strain K-12, substrain MGL655

C:Genetics:

A:Gene: phoE

A:Map position: 6 min

C:Complex: homotrimer

C:Function:

A:Description: allows passive diffusion of small hydrophilic molecules with molecular

C:Superfamily: outer membrane protein limitation

C:Keywords: homotrimer; membrane protein; porin

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-351/Product: outer membrane porin phoE #status predicted <OPP>

Query Match 61.4%; Score 1110; DB 1; Length 351;

Best Local Similarity 62.1%; Pred. No. 1.le-72;

Matches 213; Conservative 38; Mismatches 76; Indels 16; Gaps 5;

QY 1 AEIYNKDGKVDLYGKAVGLHYFSKNGENSYGGNGDWTYARLGFKGETOINSLTICYG 60

Db 22 AEIYNKDGKLDVYGVKVAHYMS-----DNASKDGQDSVIRFGFKGETQINDLTICYGR 76

QY 61 WEYFNQGNSEGAQGTGNKTRLAFAGLKYADVGSFDYGRNYGVVYDALGYTDMLPFEGG 120

Db 77 WEAEFAGNKAESDTAQ--QKTRLAFAGLKKYKDLGSGFDYGRNLGALYDVEAWTDMFPEGG 134

QY 121 D-TAYSDDFFVGRVGGVATYRNSNFFGLVDGLNFAVOYLGNKNERDRTARRSNGDGVGGGIS 179

Db 135 DSSAQTDFNTWRASGLATYRNTDFGVIDGLNLTLYQYQGNENRDYKQNGDGFSTLT 194

QY 180 YEYEG--FGIVGAYGAADRNTLQEAQPLNGKKAQEWATGLKYDANNIYLAANYGETRNA 237

Db 195 YDFGSDFAISGAYTNTDRNTNEQLQSRGTGKRAEAWATGLKYDANNIYLATFYSETRM 254

QY 238 TPTTNKFTTSGFANKTQDVLVAQYQDFGLRPSIATYTKSKAKDVEGIDVDLVNYFEV 297

Db 255 TPI-----TGGFANKTQNFVAQYQDFGLRPSLGYVLSKGKDFEGIDGLVNYIDV 308

QY 298 GATYFYNKNSYVDYIINQIDSNKLVGSDDTVAVGIVYQF 340

Db 309 GATYFYNKNSAFVDYKINQLDSNKLINNDNDIVAVGMTYQF 351

RESULT 7

AH0147

outer membrane protein C2 ompC2 [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis

C>Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001

C:Accession: AH0147

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AH0147

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-371 <CUR>

A:Cross-references: GB:AL590842; PIDN:CAC90043.1; PID:g15979264; GSPDB:GN00175

C:Genetics:

A:Gene: ompC2

C:Superfamily: outer membrane protein phoE

Query Match 59.8%; Score 1082; DB 2; Length 371;

Best Local Similarity 59.4%; Pred. No. 1.2e-70;

Matches 212; Conservative 48; Mismatches 73; Indels 24; Gaps 7;

QY 1 AEIYNKDGKNDLYGKAVGLHYFSKNGNSYGGNGDMTYARLGFKEGTQINSLDTGYGQ 60

Db 22 AEIYNKDGKNDLYGKVALHFS-----DNTKSGDKSYVRLGFGVGTQIDELSGYQ 76

QY 61 WEYFQGNNSGADAQTGNKTRLAFAGLADYGVDFGRNYGVVYDALGYTDMLPFEG 120

Db 77 WEYFNAANYAESQEAQ-DNKTFLAFAGLRYGLSGIDYGRNYGLVLYDIAAWTDMLPFCN 135

QY 121 DT-AYSDDFEVRGVGVATYRNSNFFGLVDGLNFAVQYLGKN-----ERDTARRSN 170

Db 136 DSYTRTDNFMPTGRTTGATYRNTDFGLVDGLKFLSQYQGNKGAEGETNNGRTDTSKO-N 194

QY 171 GDCVGGSSISYEY-EGFGIVGAYGAADRNTLQEAQPLNGKKAQEWATGLKYDANNIYLA 229

Db 195 GDGFGLSSEYIGAGVSGAAYASNRLTAQKNSTFGKDKADAWTTGLKYSNGVYLAA 254

QY 230 NYGETRNATPIIT-----NKFTNTSGFANKTQDVLVAQYQDFGLRPSIATYTKSKAKD 283

Db 255 NYAETRNMTPISGTAVINNVSYSGFANKTQNIELVAQYLFDFGLKPSIAYIQSGKDI 314

QY 284 EGIGDVLVNYFEVATYFYNKNSYVDYIINQIDSNKLVGSDDTVAVGIVYQF 340

Db 315 EGIGDVLVYKVIDGATYFYNKNSYVDYKINQLDNDNKLKLTNDNVVALGLVYQF 371

RESULT 8

S25520

outer membrane porin phoE precursor - Citrobacter freundii

C:Species: Citrobacter freundii

C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 20-Aug-1999

C:Accession: S25520

R:Splittings, G.; Ockhuysen, C.; Hofstra, H.; Tommassen, J.

submitted to the EMBL Data Library, August 1992

A:Reference number: S25520

A:Accession: S25520

A:Molecule type: DNA

A:Residues: 1-351 <SPI>

A:Cross-references: EMBL:X68021; NID:g40474; PIDN:CAA48162.1; PID:g40475

A>Note: this is one of the proteins induced when the E. coli cells are grown under ph

d compounds, and some other negatively charged solutes

C:Genetics:

A:Gene: phoE

C:Complex: homotrimer

C:Function:

A:Description: allows passive diffusion of small hydrophilic molecules with molecular

A>Note: induced under phosphate limitation

C:Superfamily: outer membrane protein phoE

C:Keywords: homotrimer; membrane protein; porin

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-351/Product: outer membrane porin phoE #status predicted <OPP>

Query Match 59.1%; Score 1070; DB 2; Length 351;

Best Local Similarity 59.2%; Pred. No. 8.1e-70;

Matches 203; Conservative 48; Mismatches 76; Indels 16; Gaps 5;

QY 1 AEIYNKDGKNDLYGKAVGLHYFSKNGNSYGGNGDMTYARLGFKEGTQINSLDTGYGQ 60

Db 22 AEIYNKDGKNDLYGKVALHFS-----DGDQSYIRLGFKEGTQINDELTYGR 76

QY 61 WEYFQGNNSGADAQTGNKTRLAFAGLADYGVDFGRNYGVVYDALGYTDMLPFEG 120

Db 77 WEAEFAGNKAESDSNQ--OKTRLAFAGSKLNLGSLFDYGRNLGALYDVAWTDMFPFEG 134

QY 121 D-TAYSDDFEVRGVGVATYRNSNFFGLVDGLNFAVQYLGKNERTDARRSNGDVGGSIS 179

Db 135 DSSAQTDFNTWRASGLATYRNTDFGVVDGLDLTLQYQGNKQNRDVRKKQNGDGFSTYT 194

QY 180 YEYEG--FGIVGAYGAADRNTLQEAQPLNGKKAQEWATGLKYDANNIYLAANYGETRNA 237

Db 195 YDFGSDFAVSGAYTNSDRTNQNLQTRGTGDKAEAWATGLKYDANDIYATFYSETRM 254

QY 238 TPTTNKFTTSGFANKTQDVLVAQYQDFGLRPSIATYTKSKAKDVEGIDVDLVNYFEV 297

Db 255 TPI-----SGFANKTQNFVAQYQDFGLRPSLGYVLSKGKDFEGNEDLVNYIDV 308

QY 298 GATYFYNKNSYVDYIINQIDSNKLVGSDDTVAVGIVYQF 340

Db 309 GATYFYNKNSAFVDYKINQLDSNKLINNDNDIVAVGMTYQF 351

RESULT 9

S25525

outer membrane porin phoE precursor - Salmonella typhimurium

N:Alternate names: phosphate-limitation-inducible outer membrane pore protein phoE

C:Species: Salmonella typhimurium

C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 20-Aug-1999

C:Accession: S25525; JQ1937

R:Splittings, G.; Elders, R.; van Hith, B.; Hofstra, H.; Tommassen, J.

submitted to the EMBL Data Library, August 1992

A:Reference number: S25525

A:Accession: S25525

A:Molecule type: DNA

A:Residues: 1-350 <SPI>

A:Cross-references: EMBL:X68023; NID:g47821; PIDN:CAA48164.1; PID:g47822

A:Experimental source: strain LT(2)

R:Splittings, G.; Elders, R.; van Lith, B.; Hofstra, H.; Tommassen, J.

Gene 122, 45-52, 1992

A:Title: Characterization of the Salmonella typhimurium phoE gene and development of

A:Reference number: JQ1937; MUID:93083994; PMID:1280609

A:Accession: JQ1937

A:Molecule type: DNA

A:Residues: 1-197;199-350 <SP2>

A:Cross-references: EMBL:X68023

A>Note: neither the full codon nor the amino acid are given for residue 198 in Figure

C:Genetics:

A:Gene: phoE



C:Superfamily: outer membrane protein phoE  
C:Keywords: membrane protein; porin; trimer  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-330/Product: outer membrane porin phoE #status predicted <MAT>

Query Match 58.5%; Score 1058; DB 2; Length 350;  
Best Local Similarity 58.0%; Pred. No. 5.9e-69;  
Matches 199; Conservative 46; Mismatches 82; Indels 16; Gaps 5;

QY 1 AEIYNKGNKVDLYGRAVLHYFSKNGENSYGNGDMYARLGFGETQINSDLTGYGQ 60  
Db 21 AEIYNKGNKLDYGVKAMHYSYDSK-----DGDQSYVRFEGKGETQINDLTGYGR 75  
QY 61 WEYNFGNNSGADAGTGNKTRLAFLAGLVADYSGFGRNYGVVYDALGYTDMLPFEFG 120  
Db 76 WEAEFASNAKESDSSQ--QKTRLAFLAGLVADYSGFGRNYGVVYDALGYTDMLPFEFG 133  
QY 121 D-TAYSDDFVGRVGVVATYRNFFGLVDGLNFVAYGLGKNERDTRARRSNGDVGSGSIS 179  
Db 134 DSSAQTDNFMTRKASGLATYRNDFEGIVDGLDLTLQYQKNGEDRDVKKQNGDGFSTVS 193  
QY 180 YEYEG--FGIVGAYGAADRTNLOEAPLGNKKAQWATGLKYDANNIYLAANYGETRNA 237  
Db 194 YDGGSDFAVSGAYTLSDRTREQLRRGTGDKAEAWATGVKYDANDIYIATFYSETRNM 253  
QY 238 TPTTNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYTKSKAKDVEGIDVLAANYFEV 297  
Db 238 TPTTNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYTKSKAKDVEGIDVLAANYFEV 297  
Db 254 TPV-----SGGFANKTQNEAVIQYQDFGLRPSIATYTKSKAKDVEGIDVLAANYFEV 307  
QY 298 GATYYFNKNKMSVYDYIINOIDSNDKLGVSDDDTVAVGIVYQF 340  
Db 308 GATYYFNKNKMSAFDYKINQLDSNTLGINDDDIVAIGLTYQF 350

RESULT 10  
AF0543

outer membrane pore protein E precursor STY0365 [imported] - Salmonella enterica subsp.  
C:Species: Salmonella enterica subsp. enterica serovar Typh  
A:Note: this species has also been called Salmonella typhi  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AF0543  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
A:Reference number: AB0502; PMID:11677608  
A:Accession: AF0543  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-350 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD08790.1; PID:g16501606; GSPDB:GN00176  
C:Gene: STY0365  
C:Superfamily: outer membrane protein phoE

Query Match 58.5%; Score 1058; DB 2; Length 350;  
Best Local Similarity 58.0%; Pred. No. 5.9e-69;  
Matches 199; Conservative 46; Mismatches 82; Indels 16; Gaps 5;

QY 1 AEIYNKGNKVDLYGRAVLHYFSKNGENSYGNGDMYARLGFGETQINSDLTGYGQ 60  
Db 21 AEIYNKGNKLDYGVKAMHYSYDSK-----DGDQSYVRFEGKGETQINDLTGYGR 75  
QY 61 WEYNFGNNSGADAGTGNKTRLAFLAGLVADYSGFGRNYGVVYDALGYTDMLPFEFG 120  
Db 76 WEAEFASNAKESDSSQ--QKTRLAFLAGLVADYSGFGRNYGVVYDALGYTDMLPFEFG 133  
QY 121 D-TAYSDDFVGRVGVVATYRNFFGLVDGLNFVAYGLGKNERDTRARRSNGDVGSGSIS 179  
Db 134 DSSAQTDNFMTRKASGLATYRNDFEGIVDGLDLTLQYQKNGEDRDVKKQNGDGFSTVS 193

QY 180 YEYEG--FGIVGAYGAADRTNLOEAPLGNKKAQWATGLKYDANNIYLAANYGETRNA 237  
Db 194 YDGGSDFAVSGAYTLSDRTREQLRRGTGDKAEAWATGVKYDANDIYIATFYSETRNM 253  
QY 238 TPTTNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYTKSKAKDVEGIDVLAANYFEV 297  
Db 254 TPV-----SGGFANKTQNEAVIQYQDFGLRPSIATYTKSKAKDVEGIDVLAANYFEV 307  
QY 298 GATYYFNKNKMSVYDYIINOIDSNDKLGVSDDDTVAVGIVYQF 340  
Db 308 GATYYFNKNKMSAFDYKINQLDSNTLGINDDDIVAIGLTYQF 350

RESULT 11  
MMBP22

outer membrane porin lc precursor - phage PA2  
C:Species: phage PA2  
A:Note: host Escherichia coli  
C:Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 02-Jul-1998  
C:Accession: D25647  
R:Blasband, A.J.; Marcotte Jr., W.R.; Schnaitman, C.A.  
J. Biol. Chem. 261, 12723-12732, 1986  
A:Title: Structure of the lc and nmpc outer membrane porin protein genes of lambdaoid  
A:Reference number: A25647; MUID:86304457; PMID:3017988  
A:Accession: D25647  
A:Molecule type: DNA  
A:Residues: 1-365 <BLA>  
C:Genetics: lc  
C:Superfamily: outer membrane protein phoE  
C:Keywords: membrane protein; porin; trimer  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-365/Product: outer membrane porin lc #status predicted <MAT>

Query Match 58.1%; Score 1051; DB 1; Length 365;  
Best Local Similarity 59.8%; Pred. No. 2e-68;  
Matches 211; Conservative 35; Mismatches 83; Indels 24; Gaps 7;

QY 1 AEIYNKGNKVDLYGRAVLHYFSKNGENSYGNGDMYARLGFGETQINSDLTGYGQ 60  
Db 24 AEIYNKGNKLDYGVKAMHYSYDSK-----DGDQSYVRFEGKGETQINDLTGYGQ 78  
QY 61 WEYNFGNNSGADAGTGNKTRLAFLAGLVADYSGFGRNYGVVYDALGYTDMLPFEFG 120  
Db 79 WEYEFAGNRAE--SOGSSKOKTRLAFLAGLVADYSGFGRNYGVVYDALGYTDMLPFEFG 137  
QY 121 DT-AYSDDFVGRVGVVATYRNFFGLVDGLNFVAYGLGKNER--DTARRSNGDVG 176  
Db 138 DTWTQTDVMTGRTGTGATYRNNDFFGLVDGLNFVAYGLGKNERDSDFDNTEGNGDGF 197  
QY 177 SISYEYEGFICGAYGAADRTNLOE-----AQLGNGKKAQWATGLKYDANNIYLAAN 230  
Db 198 SATYEYEGFICGAYGAADRTNLOE-----AQLGNGKKAQWATGLKYDANNIYLAAN 257  
QY 231 YGTRNATPTTNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYTKSKAKDVEGIDVD 290  
Db 258 YSETQNTVFADHFV-----ANKAQNFEAVAYQYQDFGLRPSIATYTKSKAKDVEGIDVD 312  
QY 291 LVNYFEVATYRNKNNSTVDYIINOIDSNDKLGVSDDDTVAVGIVYQF 340  
Db 313 LVNYFEVATYRNKNNSTVDYIINOIDSNDKLGVSDDDTVAVGIVYQF 365

RESULT 12  
MMECP2

outer membrane porin ompC precursor - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 01-Mar-2002  
C:Accession: A18885; B25029; E64991  
R:Mizuno, T.; Chou, M.Y.; Inouye, M.  
J. Biol. Chem. 258, 6932-6940, 1983  
A:Title: A comparative study on the genes for three porins of the Escherichia coli ou  
A:Reference number: A20867; MUID:83213433; PMID:6304064

A:Accession: A20867  
A:Molecule type: DNA  
A:Residues: 1-367 <MI2>  
A:Cross-references: GB:K00541; GB:M10314; GB:M14188; NID:g147007; PIDN:AAA24243.1; PID:g147007  
R:Mizuno, T.; Chou, M.Y.; Inouye, M.  
FEBS Lett. 151, 159-164, 1983  
A:Title: DNA sequence of the promoter region of the ompC gene and the amino acid sequence of the protein  
A:Reference number: A18885  
A:Accession: A18885  
A:Molecule type: DNA  
A:Residues: 1-40 <MI2>  
R:Nogami, T.; Mizuno, T.; Mizushima, S.  
J. Bacteriol. 164, 797-801, 1985  
A:Title: Construction of a series of ompF-ompC chimeric genes by in vivo homologous recombination  
A:Reference number: A91809; MUID:86033642; PMID:2997131  
A:Accession: B25029  
A:Molecule type: DNA  
A:Residues: 32-57 <NOG>  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: E64991  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-367 <BLAT>  
A:Cross-references: GB:AE000310; GB:U00096; NID:g2367131; PIDN:AAC75275.1; PID:g1788544;  
A:Experimental source: strain K-12, substrain MG1655  
C:Genetics:  
A:Gene: ompC  
A:Map position: 47 min  
C:Function:  
A:Description: one of the E. coli major outer membrane proteins that form passive diffusion barriers  
C:Superfamily: outer membrane protein phoE  
C:Keywords: membrane protein; porin; trimer  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-367/Product: outer membrane porin ompC #status predicted <MAT>

Query Match 58.1%; Score 1051; DB 1; Length 367;  
Best Local Similarity 58.9%; Pred. No. 2e-68;  
Matches 212; Conservative 45; Mismatches 69; Indels 34; Gaps 8;

Qy 1 AEIYNKDGKVDLYGKAVGLHYFSKNGNSYGGNGDMTYARLGPKGETQINSDLTGYGQ 60  
Db 22 AEVYNKDGKVDLYGKVDGLHYFS-----DNKDVDDQDTYMLRGLPKGETQVTDQLTGYGQ 76

Qy 61 WEYNFQGNNSGADAQTGNKTRLAFAGLKYADVGSFDYGRNYYVVDALGYTDLMPFEGG 120  
Db 77 WEYQIQGNSAEN---ENNSWTRVAFAGLKFQDVGSFDYGRNYYVVDTSWTDVLPFEGG 133

Qy 121 DTAYSDDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDT-----A 166  
Db 134 DTYGSDFNFMQGRNGFATYRNTDFGLVDGLNFAVQYOGKNGNPSGEGFTSGVTNNGRDA 193

Qy 167 RNSGDDGVGGSISYEGFVGVAGVAGAADRTNLQE-ADPLGNGKKAQEWATGLKYDANNI 225  
Db 194 LRQNGDGVGGSITDYEGFVGGTGAISSSKRTDAQNTAAAYIGNGDRATYTGGLKYDANNI 253

Qy 226 YLAANYGSTRNATPTTNKFTNSGANKTQDVLVAQYQDFGLRPSIATYKSKAKDV-E 284  
Db 234 YLAAYQTTYNATRV-----GSLGWANKAQNFEAVQYQDFGLRPSLAYLQSKGNLGR 308

Qy 285 GIGVDVLYNFEVGAATYFVFNKMSYVDYIINQIDSNKLE-----GVGSDDTVAVGIVTYQF 340  
Db 309 GYDDEDLKYVDVGATYFVFNKMSYVDYKINLLD-DNQFTEDAGINTDNIVALGLVTYQF 367

RESULT 13  
S36606  
outer membrane porin phoE precursor - Salmonella typhi  
C:Species: Salmonella typhi  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 20-Aug-1999

C:Accession: S36606  
R:Torres, A.; Puente, J.L.; Calva, E.  
submitted to the EMBL Data Library, August 1993  
A:Reference number: S36606  
A:Accession: S36606  
A:Molecule type: DNA  
A:Residues: 1-350 <TOR>  
A:Cross-references: EMBL:X74595; NID:g397159; PIDN:CAA52672.1; PID:g397160  
A:Experimental source: isolate IMSS-1  
A:Note: this is one of the proteins induced when the E. coli cells are grown under phd compounds, and some other negatively charged solutes  
C:Genetics:  
A:Gene: phoE  
C:Complex: homotrimer  
C:Function:  
A:Description: allows passive diffusion of small hydrophilic molecules with molecular weight less than 600  
A:Note: induced under phosphate limitation  
C:Superfamily: outer membrane protein phoE  
C:Keywords: membrane protein; porin  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-350/Product: outer membrane porin phoE #status predicted <OPP>

Query Match 57.7%; Score 1044; DB 2; Length 350;  
Best Local Similarity 57.4%; Pred. No. 6e-68;  
Matches 197; Conservative 47; Mismatches 83; Indels 16; Gaps 5;

Qy 1 AEIYNKDGKVDLYGKAVGLHYFSKNGNSYGGNGDMTYARLGPKGETQINSDLTGYGQ 60  
Db 21 AEVYNKDGKVDLYGKVDGLHYFSKNGNSYGGNGDMTYARLGPKGETQINSDLTGYGQ 75

Qy 61 WEYNFQGNNSGADAQTGNKTRLAFAGLKYADVGSFDYGRNYYVVDALGYTDLMPFEGG 120  
Db 76 WEAEAGAKNAESDSQ--QKTRLAFAGLKYADVGSFDYGRNYYVVDALGYTDLMPFEGG 133

Qy 121 D-TAYSDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRSNGDVGGSIS 179  
Db 134 DSAQTDNFMTRKSLATYRNTDFGIVDGLDLTLQYOGKNERDVRKQNGDGFSTVS 193

Qy 180 YEYEG-FGIVGAYGAADRTNLQEAPLGNKKAQEWATGLKYDANNIYLAANYGETRNA 237  
Db 194 YDFGSDFAVSGAYTLSDRTREQLQRCTGDKAGWATGVKYDANDIYIATFYSETRM 253

Qy 238 TPTTNKFTNSGANKTQDVLVAQYQDFGLRPSIATYKSKAKDVEGIDVDLVNFEV 297  
Db 254 TPV-----SGGFANKTNFEAVIQYQDFGLRPSISLVLSKDGIEGVSDELVNYIDV 307

Qy 298 GATYFVFNKMSYVDYIINQIDSNKLEGVGSDDTVAVGIVTYQF 340  
Db 308 GAIYFVFNKMSAFVDYKINQLDSDNTLGINDDDDIIVAILTYQF 350

RESULT 14  
MNECNC  
outer membrane porin nmpC precursor - Escherichia coli (strain K-12) cryptic lambdaoid  
C:Species: Escherichia coli  
C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 01-Mar-2002  
A:Accession: A25647; S66594; G64787  
R:Blasband, A.J.; Marcotte Jr., W.R.; Schnaitman, C.A.  
J. Biol. Chem. 261, 12723-12732, 1986  
A:Title: Structure of the lc and nmpC outer membrane porin protein genes of lambdaoid  
A:Reference number: A25647; MUID:86304457; PMID:3017988  
A:Contents: mutant strain CS384  
A:Accession: A25647  
A:Molecule type: DNA  
A:Residues: 1-365 <BLA>  
R:Mandl, A.A.; Sharples, G.J.; Mandal, T.N.; Lloyd, R.G.  
J. Mol. Biol. 257, 561-573, 1996  
A:Title: Holiday junction resolvases encoded by homologous rusA genes in Escherichia coli  
A:Reference number: S66579; MUID:96196428; PMID:8648624  
A:Accession: S66594  
A:Molecule type: DNA  
A:Residues: 347-365 <MANH>  
A:Cross-references: EMBL:X92587; NID:g1051136; PIDN:CAA63325.1; PID:g1051145

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C.  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of *Escherichia coli* K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: G64787  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 'MNIYRATVSPFNSSKKGLT', 1-325, 'N', 327-347, 'EGANKSLI' <BLAT>  
A:Cross-references: GB:AE000160; GB:U00096; NID:g1786751; PIDN:AAC73654.1; PID:g1786765;  
A:Experimental source: strain K-12, substrain MG1655  
C:Comment: In wild-type strains of *E. coli* K-12, the mmpC open reading frame is interrupted  
in CS348, the ISS element has been deleted and mmpC is expressed.  
C:Genetics:  
A:Gene: mmpC  
A:Map position: 12 min  
A:Genome: cryptic lambdaoid prophage DLP12  
C:Superfamily: outer membrane protein phoE  
C:Keywords: membrane protein; porin; trimer  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-365/Product: outer membrane porin mmpC #status predicted <MAT>

Query Match 57.6%; Score 1042; DB 1; Length 365;  
Best Local Similarity 59.5%; Pred. No. 8.9e-68;  
Matches 210; Conservative 35; Mismatches 84; Indels 24; Gaps 7;

QY 1 AEIYNKDKNVDLYGKAVGLHYFSGKNGENSYGGNGDMTYARLGFKGTQINSDLTGYGQ 60  
DB 24 AEIYNKDKNVDLYGKAVGLHYFSGKNGENSYGGNGDMTYARLGFKGTQINSDLTGYGQ 78  
QY 61 WEYNFQNNSEGAQOTGNKTRLAFAGLKYADVGSFDYGRNYGVVYDALGYTMDLPEFGG 120  
DB 79 WEYEFKGNRAE-SQGSKDKTRLAFAGLKYADVGSFDYGRNYGVVYDALGYTMDLPEFGG 137  
QY 121 DT-AYSDDFFVGRVGGVATYRNSNFFGLVDGLNFAVQYLGKNER---DTARRSNGDGVGG 176  
DB 138 DTWQTQDVMTQATGATYRNNDFGLVDGLNFAVQYLGKNER---DTARRSNGDGVGG 197  
QY 177 SISYEYEGFVGVGAYGADRTNLOE-----AQPLNGKKAQWATGLKYDANNIYLAAN 230  
DB 198 SATYEYEGFVGVGAYGADRTNLOE-----AQPLNGKKAQWATGLKYDANNIYLAAN 257  
QY 231 YGETRATPTTNKFTNTSGFANTQDVLVLAQYQDFGLRPSIAYTKSKAKDVEGIGDGD 290  
DB 258 YSETQNTVTFADHFV-----ANKAQNFEAVAQYQDFGLRPSIAYTKSKAKDVEGIGDGD 312  
QY 291 LVNYFEVGATYFNKNMSTYVDYIIQIDSDN---KLGVGSDDTVAVGIVYQF 340  
DB 313 LVKYVDVGATYFNKNMSTYVDYIIQIDSDN---KLGVGSDDTVAVGIVYQF 365

RESULT 15  
D64888  
outer membrane porin bl377 precursor - *Escherichia coli* (strain K-12)  
C:Species: *Escherichia coli*  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
C:Accession: D64888  
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C.  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of *Escherichia coli* K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: D64888  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-377 <BLAT>  
A:Cross-references: GB:AE000234; GB:U00096; NID:g1787633; PIDN:AAC74459.1; PID:g1787641;  
A:Experimental source: strain K-12, substrain MG1655  
C:Superfamily: outer membrane protein phoE  
C:Keywords: membrane protein; porin; trimer  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-377/Product: probable outer membrane porin bl377 #status predicted <OPP>

Query Match 57.2%; Score 1035; DB 2; Length 377;  
Best Local Similarity 56.9%; Pred. No. 2.9e-67;  
Matches 209; Conservative 42; Mismatches 78; Indels 38; Gaps 9;

QY 1 AEIYNKDKNVDLYGKAVGLHYFSGKNGENSYGGNGDMTYARLGFKGTQINSDLTGYGQ 60  
DB 22 AEVYNKDKNVDLYGKAVGLHYFSGKNGENSYGGNGDMTYARLGFKGTQINSDLTGYGQ 76  
QY 61 WEYNFQNNSEGAQOTGNKTRLAFAGLKYADVGSFDYGRNYGVVYDALGYTMDLPEFGG 120  
DB 77 WEYNFQNNSEGAQOTGNKTRLAFAGLKYADVGSFDYGRNYGVVYDALGYTMDLPEFGG 134  
QY 121 DT-AYSDDFFVGRVGGVATYRNSNFFGLVDGLNFAVQYLGKNER---RDTPAR 167  
DB 135 DSYTNADNFWTGRANGVATYRNTDFGLVNLNFAVQYLGKNER---RDTPAR 193  
QY 168 RNSGDGVGGISYSEY-EGFGIVGAYGAADRTNLOEAPLNGKKAQWATGLKYDANNIY 226  
DB 194 HENGSGWGLSTTYDLGNGFSAGAAATSSDRTNDQVNIHTAAGDKADAWTAGLKYDANNIY 253  
QY 227 LAANYGETRATPTTNKFTNTSGFANTQDVLVLAQYQDFGLRPSIAYTKSKAKDVE-- 284  
DB 254 LATMYSETRNTPTPEGD---SDYAVANKTQNPETVAQYQDFGLRPAVSVFLMSKGRDLHAA 310  
QY 285 -----GIGDVLVNYFEVGATYFNKNMSTYVDYIIQIDSDNKL-----GVGSDDTVA 333  
DB 311 GGADNPAGVDDKOLVKYADIGATYFNKNMSTYVDYIIQIDSDNKL-----GVGSDDTVA 370  
QY 334 VGIVYQF 340  
DB 371 LGLVYQF 377

Search completed: December 18, 2002, 16:22:23  
Job time: 11.0756 secs

***This Page Blank (uspto)***

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:06:04 : Search time 5.23408 Seconds  
(without alignments)  
2702.183 Million cell updates/sec

Title: US-09-490-291-9  
Perfect score: 1809  
Sequence: 1 AEIYNKDGKLYGKAVGL.....NKLGVGSDTVAVGIVYQFA 341

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1805	99.8	362	1	OMPFC_ECOLI
2	1110	61.4	351	1	PHOE_ECOLI
3	1077.5	59.6	362	1	OMPFC_SALTY
4	1072.5	59.3	350	1	PHOE_ENTCL
5	1072	59.3	374	1	OMPFC_SERMA
6	1070	59.1	351	1	PHOE_CITFR
7	1058	58.5	350	1	PHOE_SALTY
8	1058	58.5	350	1	PHOE_SALTY
9	1051	58.1	367	1	OMPFC_ECOLI
10	1050	58.0	351	1	PHOE_KLEPN
11	1046	57.8	365	1	PORI_BPPA2
12	1042	57.6	365	1	NMPC_ECOLI
13	1041	57.5	349	1	PHOE_KLEOX
14	1035	57.2	377	1	OMPFC_ECOLI
15	1031.5	57.0	363	1	OMPFC_SALTY
16	1031.5	57.0	363	1	OMPFC_SALTY
17	1030	56.9	363	1	OMPFC_KLEPN
18	1027	56.8	383	1	OMPFC_SALTY
19	1023.5	56.6	378	1	OMPFC_SALTY
20	999	55.2	397	1	YEDS_ECOLI
21	976.5	54.0	394	1	OMP1_SALTY
22	902	49.9	342	1	OMPFC_RAHQ
23	894.5	49.4	376	1	OMPFC_SERMA
24	885	48.9	369	1	OMPFC_XENNE
25	573.5	31.7	382	1	PORL_BUCAL
26	236.5	13.1	341	1	OMPFC_VIBCH
27	218	12.1	341	1	OMPFC_PHOPR
28	170.5	9.4	353	1	OP25_HAEIN
29	151.5	8.4	371	1	OP25_HAEIN
30	150.5	8.3	331	1	OMB_NEIMB
31	141.5	7.8	360	1	OP26_HAEIN
32	139.5	7.7	359	1	OP21_HAEIN
33	139	7.7	361	1	OP29_HAEIN

34	138.5	7.7	348	1	OMB1_NEIGO
35	138.5	7.7	361	1	OP22_HAEIN
36	135.5	7.5	369	1	OP27_HAEIN
37	133	7.4	1829	1	FRPC_NEIMB
38	131.5	7.3	348	1	OMB2_NEIGO
39	131	7.2	331	1	FRPC_NEIMB
40	131	7.2	1829	1	FRPC_NEIMB
41	130	7.2	367	1	OP2A_HAEIN
42	129.5	7.2	363	1	OP2B_HAEIN
43	129	7.1	331	1	OMB1_NEIMB
44	129	7.1	331	1	OMB3_NEIMB
45	127	7.0	1310	1	VAC3_HELPY

# ALIGNMENTS

## RESULT 1

ID	OMPFC_ECOLI	STANDARD	PRT	362 AA.
AC	P02931;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	21-JUL-1986 (Rel. 01, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Outer membrane protein F precursor (Porin ompF) (Outer membrane protein 1A) (Outer membrane protein 1A) (Outer membrane protein B).			
DE	OMPFC OR TOLF OR CMLB OR COA OR CRY OR B0929.			
GN	Escherichia coli.			
OS	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Escherichia.			
OX	NCBI_TaxID=562;			
RN	[1]_TaxID=562;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12;			
RX	MEDLINE=83090452; PubMed=6294623;			
RA	Inokuchi K., Mutoh N., Matsuyama S.-I., Mizushima S.;			
RT	"Primary structure of the ompF gene that codes for a major outer membrane protein of Escherichia coli K-12."			
RL	Nucleic Acids Res. 10:6957-6968(1982).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12 / MG1655;			
RX	MEDLINE=97426617; PubMed=9278503;			
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,			
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,			
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,			
RA	Mau B., Shao Y.;			
RT	"The complete genome sequence of Escherichia coli K-12."			
RL	Science 277:1453-1474(1997).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12;			
RX	MEDLINE=97061202; PubMed=8905232;			
RA	Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,			
RA	Ikeno K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,			
RA	Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,			
RA	Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,			
RA	Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,			
RA	Yano M., Horiuchi T.;			
RT	"A 718-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map."			
RL	DNA Res. 3:137-155(1996).			
RN	[4]			
RP	SEQUENCE OF 1-37 FROM N.A.			
RX	MEDLINE=82139379; PubMed=7037455;			
RA	Mutoh N., Inokuchi K., Mizushima S.-I.;			
RT	"Amino acid sequence of the signal peptide of OmpF, a major outer membrane protein of Escherichia coli."			
RL	FEBS Lett. 137:171-174(1982).			
RN	[5]			
RP	SEQUENCE OF 23-362.			
RX	MEDLINE=82256494; PubMed=7049161;			
RA	Chen R., Kramer C., Schmidmayr W., Chen-Schmeisser U., Henning U.;			

18195	neisseria g
20149	haemophilus
Q48218	haemophilus
Q9JYV5	neisseria m
P20148	neisseria g
P30688	neisseria m
P55127	neisseria m
Q48219	haemophilus
Q48220	haemophilus
P30687	neisseria m
P30689	neisseria m
Q48253	haemophilus

RT Primary structure of major outer-membrane protein I (ompF protein,  
RL porin) of *Escherichia coli* B/r.";  
RN Biochem. J. 203:33-43(1982).  
[6]  
RP SEQUENCE OF 33-63 FROM N.A.  
RX MEDLINE-86033642; PubMed-2997131;  
RA Nogami T., Mizuno T., Mizushima S.;  
RT "Construction of a series of ompF-ompC chimeric genes by in vivo  
RT homologous recombination in *Escherichia coli* and characterization of  
RT the translational products.";  
RL J. Bacteriol. 164:797-801(1985).  
RN [7]  
RP SEQUENCE OF 23-34 AND 39-47.  
RC STRAIN-K12 / EMG2;  
RX MEDLINE-97443975; PubMed-9298646;  
RA Link A.J., Robison K., Church G.M.;  
RT "Comparing the predicted and observed properties of proteins encoded  
RT in the genome of *Escherichia coli* K-12.";  
RL Electrophoresis 18:1259-1313(1997).  
RN [8]  
RP SEQUENCE OF 23-27.  
RC STRAIN-K12 / W3110;  
RX MEDLINE-98291876; PubMed-9629924;  
RA Molloy M.P., Herbert B.R., Walsh B.J., Tyler M.I., Traini M.,  
RA Molloy M.P., Herbert B.R., Walsh B.J., Tyler M.I., Traini M.,  
RA Sanchez J.-C., Hochstrasser D.F., Williams K.L., Gooley A.A.;  
RT "Extraction of membrane proteins by differential solubilization for  
RT separation using two-dimensional gel electrophoresis.";  
RL Electrophoresis 19:837-844(1998).  
RN [9]  
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).  
RX MEDLINE-92375189; PubMed-1380671;  
RA Cowan S.W., Schirmer T., Rummel G., Steiert M., Ghosh R.,  
RA Pauptit R.A., Jansonius J.N., Rosenbusch J.P.;  
RT "Crystal structures explain functional properties of two *E. coli*  
RT porins.";  
RL Nature 358:727-733(1992).  
RN [10]  
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) OF MUTANT ASP-141.  
RX MEDLINE-95024177; PubMed-7524100;  
RA Jeanteur D., Schirmer T., Fourle D., Simonet V., Rummel G., Widmer C.,  
RA Rosenbusch J.P., Pattus F., Pages J.M.;  
RT "Structural and functional alterations of a colicin-resistant mutant  
RT of OmpF porin from *Escherichia coli*.";  
RL Proc. Natl. Acad. Sci. U.S.A. 91:10675-10679(1994).  
RN [11]  
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).  
RX MEDLINE-99060042; PubMed-9843370;  
RA Phale P.S., Philippsen A., Kiefhaber T., Koebnik R., Phale V.P.,  
RA Schirmer T., Rosenbusch J.P.;  
RT "Stability of trimeric OmpF porin: the contributions of the latching  
RT loop L2.";  
RL Biochemistry 37:15663-15670(1998).  
CC -!- FUNCTION: OMPF IS A PORIN THAT PASSIVE DIFFUSION PORES WHICH ALLOW  
CC SMALL MOLECULAR WEIGHT HYDROPHILIC MATERIALS ACROSS THE OUTER  
CC MEMBRANE. IT IS ALSO RECEPTOR FOR THE BACTERIOPHAGE T2.  
CC -!- SUBUNIT: HOMOTRIMER.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.  
CC -!- SIMILARITY: BELONGS TO THE OMP/PHOE FAMILY OF PORINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; J01655; AAA2444.1; -;  
DR EMBL; AE000195; AAC74015.1; -;  
DR EMBL; D90730; BAA35675.1; -;  
DR EMBL; D90731; BAA35681.1; -;  
DR PIR; A03431; MMECF.  
DR PIR; A25029; A25029.

DR PDB; 2OMF; 07-DEC-95.  
DR PDB; 1OPF; 07-FEB-95.  
DR PDB; 1GFM; 07-DEC-96.  
DR PDB; 1GFN; 07-DEC-96.  
DR PDB; 1GFO; 07-DEC-96.  
DR PDB; 1GFP; 07-DEC-96.  
DR PDB; 1GFQ; 07-DEC-96.  
DR PDB; 1IMP; 07-FEB-95.  
DR PDB; 1BT9; 13-JAN-99.  
DR SWISS-2DPAGE; P02931; COLI.  
DR ECO2DBASE; B036.0; 6TH EDITION.  
DR EcoGene; EG10671; ompF.  
DR InterPro; IPR003229; OMP\_2.  
DR InterPro; IPR001702; Porin\_gram-ve.  
DR Pfam; PF00267; Gram-ve\_porins; 1.  
DR PRINTS; PR00182; ECOLNEIPORIN.  
DR PROSITE; PS00576; GRAM\_NEG\_PORIN; 1.  
KW Outer membrane; Transmembrane; Porin; Phage recognition; Signal;  
KW 3D-structure; Complete proteome.  
FT SIGNAL 1 22  
FT CHAIN 23 362 OUTER MEMBRANE PROTEIN F.  
FT CONFLICT 88 88 Q -> E (IN REF. 5).  
FT CONFLICT 139 139 E -> G (IN REF. 5).  
FT CONFLICT 284 284 Q -> L (IN REF. 5).  
FT STRAND 24 28  
FT TURN 29 30  
FT STRAND 31 45  
FT TURN 50 51  
FT STRAND 53 53  
FT STRAND 58 59  
FT STRAND 62 72  
FT STRAND 77 88  
FT TURN 95 100  
FT STRAND 102 112  
FT TURN 113 115  
FT STRAND 116 124  
FT TURN 126 127  
FT HELIX 128 131  
FT TURN 132 132  
FT TURN 143 144  
FT TURN 147 148  
FT TURN 150 151  
FT STRAND 154 163  
FT HELIX 165 169  
FT TURN 171 172  
FT STRAND 173 180  
FT STRAND 183 183  
FT TURN 188 190  
FT STRAND 192 192  
FT TURN 205 206  
FT STRAND 207 217  
FT HELIX 220 223  
FT TURN 224 224  
FT STRAND 227 227  
FT STRAND 232 244  
FT TURN 245 246  
FT STRAND 247 257  
FT STRAND 261 264  
FT TURN 265 268  
FT STRAND 269 272  
FT STRAND 275 285  
FT TURN 288 289  
FT STRAND 291 305  
FT TURN 306 308  
FT STRAND 309 324  
FT STRAND 329 338  
FT STRAND 351 351  
FT STRAND 353 361  
SQ SEQUENCE 362 AA; 39333 MW; 3F0974D96DB65464 CRC64;  
Query Match 99.8%; Score 1805; DB 1; Length 362;  
Best Local Similarity 100.0%; Pred. No. 2.9e-123;

```
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AEIYNKGNKVDLYGKAVGLHYHSKNGENS YGGNGDMTYARLGFKGTQINSDLTGYG 60
DB 23 AEIYNKGNKVDLYGKAVGLHYHSKNGENS YGGNGDMTYARLGFKGTQINSDLTGYG 82
QY 61 WEYNFQGNNEGADQATGKTRLAFAAGLKADYVGSFDFGRNYSVYDALGYTDMLEPFGG 120
DB 83 WEYNFQGNNEGADQATGKTRLAFAAGLKADYVGSFDFGRNYSVYDALGYTDMLEPFGG 142
QY 121 DTAYSDDFVGRVGGVATYRNSNFFGLVDGLNFAVQYLGKNERDTRRSNGDVGGSISY 180
DB 143 DTAYSDDFVGRVGGVATYRNSNFFGLVDGLNFAVQYLGKNERDTRRSNGDVGGSISY 202
QY 181 EYEGFGIVGAYGAADRTNLOEAQPLNGKKAQWATGLKTDANNIYLAANYGETRNATPI 240
DB 203 EYEGFGIVGAYGAADRTNLOEAQPLNGKKAQWATGLKTDANNIYLAANYGETRNATPI 262
QY 241 TNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYKSKAKDVEGIDVDLVNVEFGAT 300
DB 263 TNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYKSKAKDVEGIDVDLVNVEFGAT 322
QY 301 YFNKNNSTVYDIINOIDSNDKLGVSDDTVAVGIYQF 340
DB 323 YFNKNNSTVYDIINOIDSNDKLGVSDDTVAVGIYQF 362

RESULT 2
PHOE_ECOLI
ID PHOE_ECOLI STANDARD; PRT; 351 AA.
AC P02932;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane pore protein E precursor.
GN PHOE OR OMPE OR B0241.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=83189086; PubMed=6341601;
RA Overbeek N., Bergmans H., van Mansfeld F., Lugtenberg B.;
RT "Complete nucleotide sequence of phoE, the structural gene for the
RT phosphate limitation inducible outer membrane pore protein of
RT Escherichia coli K12."
RL J. Mol. Biol. 163:513-532(1983).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MGI655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RA Takemoto K., Mori H., Murayama N., Kataoka K., Yano M., Itoh T.,
RA Yamamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S.,
RA Mizuno T., Makino K., Nakata A., Yura T., Sempel G., Mizobuchi K.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of the
RT 4.0 - 6.0 min (189,987 - 281,416bp) region."
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
```

```
RA Davis R.W.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-22 FROM N.A.
RX MEDLINE=84297232; PubMed=6089111;
RA Deutch A.H., Rushlow K.E., Smith C.J.;
RT "Analysis of the Escherichia coli proBA locus by DNA and protein
RT sequencing."
RL Nucleic Acids Res. 12:6337-6355(1984).
RN [6]
RP MUTAGENESIS OF PHE-351.
RX MEDLINE=91162638; PubMed=1848301;
RA Struyve M., Moons M., Tommassen J.;
RT "Carboxy-terminal phenylalanine is essential for the correct assembly
RT of a bacterial outer membrane protein."
RL J. Mol. Biol. 218:141-148(1991).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (6.0 ANGSTROMS).
RX MEDLINE=91172301; PubMed=1848682;
RA Jap B.K., Walian P.J., Gehring K.;
RT "Structural architecture of an outer membrane channel as determined
RT by electron crystallography."
RL Nature 350:167-170(1991).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RX MEDLINE=92375189; PubMed=1380671;
RA Cowan S.W., Schirmer T., Rummel G., Steiert M., Ghosh R.,
RA Paupit R.A., Jansonius J.N., Rosenbusch J.P.;
RT "Crystal structures explain functional properties of two E. coli
RT porins."
RL Nature 358:727-733(1992).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RX MEDLINE=93172954; PubMed=7679770;
RA Struyve M., Visser J., Adriense H., Benz R., Tommassen J.;
RT "Topology of phoE porin: the 'eyelet' region."
RL Mol. Microbiol. 7:131-140(1993).
CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN E. COLI CELLS
CC ARE GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS
CC PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE,
CC PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
CC SOLUTES.
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: V00316; CAA23605.1;
CC EMBL: A8000132; AAC73345.1;
CC EMBL: D83536; BAA77910.1;
CC EMBL: U70214; AAB08661.1;
CC EMBL: X00786; CAA25362.1;
CC PIR: A03432; MMECEP.
CC PDB: 1PHO; 3I-OCT-93.
CC SWISS-2DPAGE: P02932; COLI.
CC ECO2DBASE: B037.0; 6TH EDITION.
CC EcoGene: EG10729; phoE.
CC InterPro: IPR003229; OMP.2.
CC InterPro: IPR001702; Porin-gram-ve.
CC Pfam: PF00267; Gram-ve-porins; 1.
CC PRINTS: PR00182; ECOLNEIPORIN.
CC PROSITE: PS00576; GRAM_NEG_PORIN; 1.
CC Outer membrane; Transmembrane; Porin; Signal; 3D-structure;
KW Complete proteome.
FT SIGNAL 1 21
FT CHAIN 22 351 OUTER MEMBRANE PORE PROTEIN E.
```

```
FT MUTAGEN 351 351 F->Y,N,S,V: LESS RESISTANT TO TRYPSIN.
FT STRAND 25 26
FT STRAND 30 35
FT STRAND 37 45
FT STRAND 52 53
FT STRAND 56 65
FT STRAND 71 82
FT TURN 87 88
FT STRAND 93 104
FT TURN 105 107
FT STRAND 108 116
FT TURN 119 119
FT HELIX 120 123
FT TURN 124 126
FT TURN 140 141
FT TURN 143 145
FT STRAND 147 157
FT TURN 159 162
FT TURN 164 165
FT STRAND 166 173
FT STRAND 176 176
FT TURN 182 183
FT STRAND 185 185
FT STRAND 188 197
FT TURN 198 201
FT STRAND 202 212
FT HELIX 215 217
FT TURN 218 219
FT STRAND 227 239
FT TURN 240 241
FT STRAND 242 252
FT STRAND 264 274
FT STRAND 280 292
FT STRAND 294 294
FT TURN 295 297
FT STRAND 298 298
FT STRAND 301 313
FT STRAND 318 327
FT TURN 334 335
FT STRAND 342 350
SQ SEQUENCE 351 AA; 249E2E362C030C9A CRC64;

Query Match 61.4%; Score 1110; DB 1; Length 351;
Best Local Similarity 62.1%; Pred. No. 3.5e-73;
Matches 213; Conservative 38; Mismatches 76; Indels 16; Gaps 5;

QY 1 AEIYNKDGKNDLYGKAVGLHYFSKNGENSYGGNGDMTYARLGFPGQTOINSDLTGYGO 60
DB 22 AEIYNKDGKNDLYGKAVGLHYFSKNGENSYGGNGDMTYARLGFPGQTOINSDLTGYGO 76

QY 61 WEYNFQGN--NSEGADQTKNTRLAFAGLKYADVGSFDYGRNYYGVVYDALGYTDMLEPFGG 120
DB 77 WEAEFAGNKAESDTAQ--QKTRLAFAGLKYKDLGSFDYGRNLGALYDVEAWTDMFPEFGG 134

QY 121 D-TATSDDFEYRGVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARNSGDGVGGSIS 179
DB 135 DSSAQTDNFMTRKASGLATYRTDFGVIDGLNLTQYQGNENRDVKKQNGDGFCTSLT 194

QY 180 YEYEG--FGIVGACAAADRNLQEAQPLGNGKKAQEWATGLKYDANNLYLAANYGETRNA 237
DB 195 YDFGSDFAISGAYNNSRTNQNQLQSGTGKRAAWATGLKYDANNLYLATFYSETRRM 254

QY 238 TPITNKTNTSGFANKTDVLVAQYQDFGLRPSIATYTKSKADVEGIDVDLVNYPEV 297
DB 255 TPI-----TGFANKTQNFEEVAQYQDFGLRPSLGSVLSKGDIEGIDEDLYNIDV 308

QY 298 GATYFENKMSYVDYIINQISDKNLGLVGSDDTVAVGIVTQF 340
DB 309 GATYFENKMSAFVDYKINQLDSNDKNLNNDDIYAVGMTYQF 351
```

RESULT 3  
OMPD\_SALTY

```
ID AC P37592;
DT 01-OCT-1994 (Rel. 30, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Outer membrane porin protein ompd precursor.
GN OMPD OR NMPC OR STM1572.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RC MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
RN [2]
RP SEQUENCE OF 80-362 FROM N.A.
RC STRAIN=SLI303;
RX MEDLINE=95011654; PubMed=7926834;
RA Hongo E., Morimyo M., Mita K., Machida I., Hama-Inaba H., Tsuji H.,
RA Ichimura S., Noda Y.;
RT "The methyl viologen-resistance-encoding gene smvA of Salmonella
RT typhimurium."
RL Gene 148:173-174(1994).
RN [3]
RP IDENTIFICATION AS OMPD.
RA Singh S.P., Miller S., Williams Y.U., Rudd K.E., Nikaido H.;
RL Unpublished observations (FEB-1996).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -!- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; AE008768; AAL20490.1; -.
CC EMBL; D26057; BAA05056.1; -.
CC HSSP; P02931; IGFN.
CC StyGene; SG10249; ompd.
CC InterPro; IPR003229; OMP_2.
CC InterPro; IPR001702; Porin_gram-ve.
CC Pfam; PF00267; Gram-ve_porins; 2.
CC PROSITE; PS00576; GRAM_NEG_PORIN; 1.
CC Outer membrane; Transmembrane; Porin; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 362 OUTER MEMBRANE PORIN PROTEIN OMPD.
FT CONFLICT 157 157 N -> K (IN REF. 2).
FT CONFLICT 258 258 MISSING (IN REF. 2).
SQ SEQUENCE 362 AA; 39679 MW; 9F68A9A7B948174C CRC64;

Query Match 59.6%; Score 1077.5; DB 1; Length 362;
Best Local Similarity 62.3%; Pred. No. 8e-71;
Matches 218; Conservative 31; Mismatches 82; Indels 19; Gaps 7;

QY 1 AEIYNKDGKNDLYGKAVGLHYFSKNGENSYGGNGDMTYARLGFPGQTOINSDLTGYGO 60
DB 22 AEIYNKDGKNDLYGKAVGLHYFSKNGENSYGGNGDMTYARLGFPGQTOINSDLTGYGO 76

QY 61 WEYNFQGN--NSEGADQTKNTRLAFAGLKYADVGSFDYGRNYYGVVYDALGYTDMLEPFGG 118
DB 77 WEYEFKGRNRTESQGADK---DKTRLAFAGLKFADYGSFDYGRNYYGVAYDYGATVDLPEFG 133
```



```

QY 119 GGDY-AYSDFFYGRVGGVATYRNSNFFGLVDGLNFAVOYLGNKRDRTARRSNGDGVGGS 177
DB 134 GDTWTQTDVFTVGTGATYNTDFFGLVEGLNFAAQYQGNKDRDAGYESNGDGFGLS 193
QY 178 ISYIEYEGFIVGAYGAADRTNLQEAQPLGN----GKKAQWATGLKYDANNIYLAANYGE 233
DB 194 ATVEYEGFVGGAAYKSDRTN-NOVKAASNLNAAGNAEWAAGLAGYDANNIYLAANYSE 252
QY 234 TRNATPITNKTNTSGFANKTQDVLVLAQYQDFGLRPSIATYKSAKOVGEGDVLVN 293
DB 253 TLMNTTFGDEAAGDAFIANKTQFEVAQYQDFGLRPSIATYKSAKOVGEGDVLVN 312
QY 294 YFVGATYFKNMSTYVDYIINOISDN---KLGVGSDDTVAAGVIYQF 340
DB 313 YIDVGATYFKNMSTYVDYIINOISDN---KLGVGSDDTVAAGVIYQF 362

RESULT 4
PHOE_ENTCL STANDARD; PRT; 350 AA.
AC 047490;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE Outer membrane pore protein E precursor.
GN PHOE.
OS Enterobacter cloacae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Enterobacter.
OX NCBI_TaxID=550;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87190417; PubMed=3032618;
RA van der Ley P., Bekkers A., van Meersbergen J., Tommassen J.;
RT "A comparative study on the phoE genes of three enterobacterial
RT species. Implications for structure-function relationships in a
RT pore-forming protein of the outer membrane.";
RL Eur. J. Biochem. 164:469-475(1987).
CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN CELLS ARE
CC GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS
CC PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE,
CC PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
CC SOLUTES.
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DB EMBL: M28296; AAA24809.1;
DR HSSP: P02932; LPHO.
DR InterPro: IPR003229; OMP_2.
DR InterPro: IPR001702; Porin_gram-ve.
DR Pfam: PF00267; Gram-ve_porins; 1.
DR PRINTS: PR00182; ECOLNEIPORIN.
DR PROSITE: PS00376; GRAM_NEG_PORIN; 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 21 BY SIMILARITY.
FT CHAIN 22 350 OUTER MEMBRANE PORE PROTEIN E.
SQ SEQUENCE 350 AA; 38511 MW; 101665FECC3AFAA CRC64;

Query Match 59.3%; Score 1072.5; DB 1; Length 350;
Best Local Similarity 60.3%; Pred. No. 1.8e-70;
Matches 207; Conservative 40; Mismatches 79; Indels 17; Gaps 5;

QY 1 AEIYNKGNKVDLYGKAVGLHYFKSGNGENSYGGNGDMYIARLFGKGETQINSDLTYGYQ 60

```

```

DB 22 AEYNNKNGKLDVYGVKAMHYISDDDTK-----DGDQTVVRFGKGETQINOLTYGYR 76
QY 61 WEYNFQGNNSGADAGTGNKTRLAFLAGLYADVGSFEDYGRNYGVYVDALGYTDMLPFEGG 120
DB 77 WEAEFAGNAKESDSSQ---KTRLAFLAGLYADVGSFEDYGRNYGVYVDALGYTDMLPFEGG 133
QY 121 D-TAYSDDFVGRVGGVATYRNSNFFGLVDGLNFAVOYLGNKRDRTARRSNGDGVGGSIS 179
DB 134 DSSAQTDNFTWKTRASGLATYRNTDFFGAIDGLDMLTQYQGNKDRDAGYESNGDGFGLSLT 193
QY 180 YEYEG--FGIVGAYGAADRTNLQEAQPLGNKKAQWATGLKYDANNIYLAANYGETRNA 237
DB 194 YDFGGTDFAVSGAYTNSDRINAQNLARAGQKAEAWATGLKYDANDIYLAANYSETRNM 253
QY 238 TPITNKTNTSGFANKTQDVLVLAQYQDFGLRPSIATYKSAKOVGEGDVLVNFEY 297
DB 254 TPI-----SGGFANKAQNPEVVAQYQDFGLRPSIATYKSAKOVGEGDVLVNFEY 307
QY 298 GATYFKNMSTYVDYIINOISDNKLGVGSDDTVAAGVIYQF 340
DB 308 GATYFKNMSTYVDYIINOISDNKLGVGSDDTVAAGVIYQF 350

RESULT 5
OMPF_SERMA STANDARD; PRT; 374 AA.
AC O33980;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE Outer membrane protein F precursor (Porin ompF).
GN OMPF.
OS Serratia marcescens.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Serratia.
OX NCBI_TaxID=615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UCC-51;
RX MEDLINE=97419518; PubMed=9274033;
RA Huttsul J.A.M., Worobec E.A.;
RT "Molecular characterization of the Serratia marcescens OmpF porin,
RT and analysis of S. marcescens OmpF and OmpC osmoregulation.";
RL Microbiology 143:2797-2806(1997).
CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DB EMBL: U01967; AAB69103.1; ALT_INIT.
DR HSSP: Q48473; IOSM.
DR InterPro: IPR003229; OMP_2.
DR InterPro: IPR001702; Porin_gram-ve.
DR Pfam: PF00267; Gram-ve_porins; 1.
DR PRINTS: PR00182; ECOLNEIPORIN.
DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 374 OUTER MEMBRANE PROTEIN F.
SQ SEQUENCE 374 AA; 41184 MW; BF821D385E441049 CRC64;

Query Match 59.3%; Score 1072; DB 1; Length 374;
Best Local Similarity 60.1%; Pred. No. 2.1e-70;
Matches 215; Conservative 38; Mismatches 81; Indels 24; Gaps 10;

```

```
Qy 1 AEIYNKGNKVDLYOKAVGLHYFSKNGENSYGGNDMTYARLGFPGKETQINSIDLTYGYQ 60
Db 23 AEIYNKGNKLDLYKGVLDYFYSKDK-----GNDQDTYVRFPGKETITDQLTYGYQ 77
Qy 61 WEYNFQGNSSGADAOCTGNKTRLAFAGLYADVGSDYGRNYYGVVYDYGALGYDMLPERGG 120
Db 78 WEYNVQSNHASEQGTG-CTKTRLGFAGLKFADYGSDFYGRNYYGVLDYVEGWTDMLEPERGG 136
Qy 121 DT-AYSDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERD--TARRSNGDGVGGS 177
Db 137 DTYYSDNFMTRGVATYRNNFFGLVDGLNFAVQYLGKNERD--TARRSNGDGVGGS 196
Qy 178 ISYEF-EGFGLVGVAGAADRTNLQBAQPLGNGKKAQWATGLKYDANNIYLAANYGETRN 236
Db 197 STYDIGEVSFGAAYASSNRDQKLRNERNRGRKADATVGAKYDANNVYLAANYAETRN 256
Qy 237 ATP1-TNKTFTN-----SGFANKTQDVLVLAQYQDFGLRPSIATYKSKAK--DVSGI 286
Db 257 MTPFGGGNFTNCAATENCGGFASKTQNFVTAQYQDFGLRPSIATYKSKAK--DVSGI 316
Qy 287 G-DVDLVNVEFGVATYRNNKSTVVDYIINQIDSDN---KLGVGSDDTAVAGIYQF 340
Db 317 GSDQDLVKYVSGVITYYNNKSTVVDYIINQIDSDN---KLGVGSDDTAVAGIYQF 374

RESULT 6
PHOE_CITFR
ID PHOE_CITFR STANDARD; PRT; 351 AA.
AC Q01605;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Outer membrane pore protein E precursor.
GN PHOE.
OS Citrobacter freundli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Citrobacter.
OX NCBI_TaxID=546;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93146376; PubMed=1337052;
RA Sierlings G., Ockhuijsen C., Hofstra H., Tommassen J.;
RT "Characterization of the Citrobacter freundlii phoe gene and
RL FEMS Microbiol. Lett. 78:199-204(1992).
CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN CELLS ARE
CC GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS
CC PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE,
CC PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
CC SOLUTES.
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; X68021; CAA48162.1; -
CC PIR; S25520; S25520.
CC DR HSP; P02932; 1PHO.
CC DR InterPro; IPR003229; OMP_2.
CC DR InterPro; IPR001702; Porin_gram-ve.
CC DR Pfam; PF00267; Gram-ve_porins; 1.
CC DR PRINTS; PR00182; ECOLNEIPORIN.
CC DR PROSITE; PS00576; GRAM_NEG_PORIN; 1.
CC KW Outer membrane; Transmembrane; Porin; Signal.
CC FT SIGNAL 1 21
CC CHAIN 22 351 OUTER MEMBRANE PORE PROTEIN E.
```

```
SO SEQUENCE 351 AA; 38807 MW; 4BE9C144070381B5A CRC64;
Query Match 59.1%; Score 1070; DB 1; Length 351;
Best Local Similarity 59.2%; Pred. No. 2.7e-70;
Matches 203; Conservative 48; Mismatches 76; Indels 16; Gaps 5;

Qy 1 AEIYNKGNKVDLYOKAVGLHYFSKNGENSYGGNDMTYARLGFPGKETQINSIDLTYGYQ 60
Db 22 AEIYNKGNKLDLYKGVKAMHYMTDYSK-----DGQSYIRLGFPGKETQINDELTYGYR 76
Qy 61 WEYNFQGNSSGADAOCTGNKTRLAFAGLYADVGSDYGRNYYGVVYDYGALGYDMLPERGG 120
Db 77 WEAEFAGNAKESDSNQ--QKTRLAFAGSKLKNLGSDFYGRNLYDLYEAWTDMFPEFG 134
Qy 121 D-TAYSDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDARRSNGDGVGGSIS 179
Db 135 DSSAQTDNFMTRKASGLATYRNTDFGVVDGLDLTLQYQGNQDRDVKKQNGDGFSTVT 194
Qy 180 YEYEG--FGIVGAYGAADRTNLQEAQPLGNGKKAQWATGLKYDANNIYLAANYGETRNA 237
Db 195 YDFGSGDFAVSGAVTNSDRTNQNLQTRGTGDKAEAWATGLKYDANDIYIATFYSETRNM 254
Qy 238 TPIITNKFTNSGFANKTQDVLVLAQYQDFGLRPSIATYKSKAKDVEGIDVDLVNVEFV 297
Db 255 TPI-----SGFANKTQNFVTAQYQDFGLRPSIATYKSKAKDVEGIDVDLVNVEFV 308
Qy 298 GATYFNKMNSTVVDYIINQIDSDNKLGVGSDDTAVAGIYQF 340
Db 309 GATYFNKMNSTVVDYIINQIDSDNKLGVGSDDTAVAGIYQF 351

RESULT 7
PHOE_SALTI
ID PHOE_SALTI STANDARD; PRT; 350 AA.
AC Q56119;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Outer membrane pore protein E precursor.
GN PHOE OR STY0365.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=IMSS-1;
RA Torres A., Puente J.L., Calva E.;
RL Submitted (AUG-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RL enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN CELLS ARE
CC GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS
CC PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE,
CC PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
CC SOLUTES.
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
CC -----
```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X74595; CAA52672.1; -;  
EMBL; AL627266; CAD08790.1; -;  
HSP; P02932; IPHO.  
InterPro; IPR003229; OMP\_2.  
InterPro; IPR001702; Porin-gram-ve.  
Pfam; PF00267; Gram-ve porins; 1.  
PRINTS; PR00182; ECOLNEIPORIN.  
PROSITE; PS00576; GRAM\_NEG\_PORIN; 1.  
Outer membrane; Transmembrane; Porin; Signal; Complete proteome.  
FT SIGNAL 1 20 BY SIMILARITY.  
FT CHAIN 21 350 OUTER MEMBRANE PORE PROTEIN E.  
FT CONFLICT 63 63 E -> K (IN REF. 1).  
FT CONFLICT 80 80 F -> G (IN REF. 1).  
FT CONFLICT 229 229 A -> G (IN REF. 1).  
SQ SEQUENCE 350 AA; 38744 MW; 890F5F8C5EF0C77 CRC64;

Query Match 58.5%; Score 1058; DB 1; Length 350;  
Best Local Similarity 58.0%; Pred. No. 2e-69;  
Matches 199; Conservative 46; Mismatches 82; Indels 16; Gaps 5;

QY 1 AEIYKDGKNDVLYGKAVGLHYFSKNGENSGYGGNDMTYARLFGKGTQINSDLTGYGQ 60  
Db 21 AEYVKNKNGKLDVYGVKAMHYMSDYDSK-----DGQSVYVRFKGTQINDQLTGYGR 75  
QY 61 WEYNFQGNSEGADAGTGNKTRLAFAGLVADYGVSGFYGRNYGVVYDALGYTDMLPFEGG 120  
Db 76 WEAEFAGNKAESDSQ--QKTRLAFAGLVADYGVSGFYGRNYGVVYDALGYTDMLPFEGG 133  
QY 121 D-TAYSDDFVGRVGGVATYRNSNFFGLVDGLNFAVYGLKNGEDTARRSGDVGGSIS 179  
Db 134 DSSAQTDNFTKRSGLATYRNTDFGIVDGLDLTLQYQGNEDRVKKGDFGFTSVS 193  
QY 180 YEYEG--FGIVGAGAADRTNLQEAQPLGNGKKAQWATGLKYDANNIYLAANYGETRNA 237  
Db 194 YDFGGSDFAVSGAYTLDSTRQNLQRGTGDKAEAWATGVKYDANDIYIATFYSETRNM 253  
QY 238 TPITNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYKSKAKDVEIGDVLVNYFEV 297  
Db 254 TPV-----SGGFANKTQNEAVIQYQDFGLRPSLGLVLSKGDIEGVGSDLVNVIDY 307  
QY 298 GATYFENKNSVYDYIIQINDSKNLGVSDDTAVAGIVYQF 340  
Db 308 GATYFENKNSAFVDYKINQLDSNTLGINDDIIVAGLTYQF 350

RESULT 8  
PHOE\_SALTY STANDARD; PRT; 350 AA.  
AC P30705;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Outer membrane pore protein E precursor.  
GN PHOE OR STM0320  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LT2;  
RX MEDLINE=93083994; PubMed=1280609;  
RA Spierings G., Elders R., van Lith B., Hofstra H., Tommassen J.;  
RT "Characterization of the Salmonella typhimurium phoE gene and development of Salmonella-specific DNA probes.";

Gene 122:45-52(1992).  
[2]  
SEQUENCE FROM N.A.  
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium LT2.";  
RL Nature 413:852-856(2001).  
CC 1- FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN CELLS ARE GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE, PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED SOLUTES.  
CC 1- SUBUNIT: HOMOTRIMER.  
CC 1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.  
CC 1- SIMILARITY: BELONGS TO THE OMP/PHOE FAMILY OF PORINS.  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X68023; CAA48164.1; -;  
EMBL; AE008709; AAL19276.1; -;  
PIR; S25525; S25525.  
HSP; P02932; IPHO.  
DR StyGene; SG10291; phoE.  
DR InterPro; IPR003229; OMP\_2.  
DR Pfam; IPR001702; Porin-gram-ve.  
DR PRINTS; PR00182; ECOLNEIPORIN.  
DR PROSITE; PS00576; GRAM\_NEG\_PORIN; 1.  
KW Outer membrane; Transmembrane; Porin; Signal; Complete proteome.  
FT SIGNAL 1 20  
FT CHAIN 21 350 OUTER MEMBRANE PORE PROTEIN E.  
SQ SEQUENCE 350 AA; 38762 MW; 9D1E3355AF59877 CRC64;

Query Match 58.5%; Score 1058; DB 1; Length 350;  
Best Local Similarity 58.0%; Pred. No. 2e-69;  
Matches 199; Conservative 46; Mismatches 82; Indels 16; Gaps 5;

QY 1 AEIYKDGKNDVLYGKAVGLHYFSKNGENSGYGGNDMTYARLFGKGTQINSDLTGYGQ 60  
Db 21 AEYVKNKNGKLDVYGVKAMHYMSDYDSK-----DGQSVYVRFKGTQINDQLTGYGR 75  
QY 61 WEYNFQGNSEGADAGTGNKTRLAFAGLVADYGVSGFYGRNYGVVYDALGYTDMLPFEGG 120  
Db 76 WEAEFAGNKAESDSQ--QKTRLAFAGLVADYGVSGFYGRNYGVVYDALGYTDMLPFEGG 133  
QY 121 D-TAYSDDFVGRVGGVATYRNSNFFGLVDGLNFAVYGLKNGEDTARRSGDVGGSIS 179  
Db 134 DSSAQTDNFTKRSGLATYRNTDFGIVDGLDLTLQYQGNEDRVKKGDFGFTSVS 193  
QY 180 YEYEG--FGIVGAGAADRTNLQEAQPLGNGKKAQWATGLKYDANNIYLAANYGETRNA 237  
Db 194 YDFGGSDFAVSGAYTLDSTRQNLQRGTGDKAEAWATGVKYDANDIYIATFYSETRNM 253  
QY 238 TPITNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYKSKAKDVEIGDVLVNYFEV 297  
Db 254 TPV-----SGGFANKTQNEAVIQYQDFGLRPSLGLVLSKGDIEGVGSDLVNVIDY 307  
QY 298 GATYFENKNSVYDYIIQINDSKNLGVSDDTAVAGIVYQF 340  
Db 308 GATYFENKNSAFVDYKINQLDSNTLGINDDIIVAGLTYQF 350

```

RESULT 9
OMPC_ECOLI
ID OMPC_ECOLI STANDARD; PRT; 367 AA.
AC P06996;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein C precursor (Porin ompC) (Outer membrane
DE protein 1B).
GN OMPC OR MEOA OR PAR OR B2215.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=83213433; PubMed=6304064;
RA Mizuno T., Chou M.-Y., Inouye M.;
RT "A comparative study on the genes for three porins of the Escherichia
RT coli outer membrane. DNA sequence of the osmoregulated ompC gene.";
RN J. Biol. Chem. 258:6932-6940(1983).
[2]
RN SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RN Science 277:1453-1474(1997).
[3]
RN SEQUENCE FROM N.A.
RX STRAIN-K12;
RA MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
RA Saito N., Sampei G., Seki Y., Sivasubraman S., Tagami H.,
RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RN DNA Res. 3:379-392(1996).
[4]
RN SEQUENCE OF 218-367 FROM N.A.
RX STRAIN-K12 / BHB2600;
RA Richterich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,
RA Church G.M.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
[5]
RN SEQUENCE OF 1-22 FROM N.A., AND SEQUENCE OF 22-40.
RX MEDLINE=83132326; PubMed=6297988;
RA Mizuno T., Chou M.-Y., Inouye M.;
RT "DNA sequence of the promoter region of the ompC gene and the amino
RT acid sequence of the signal peptide of pro-OmpC protein of
RT Escherichia coli.";
RN FEBS Lett. 151:159-164(1983).
[6]
RN SEQUENCE OF 32-57 FROM N.A.
RX MEDLINE=86033642; PubMed=2997131;
RA Nogami T., Mizuno T., Mizushima S.;
RT "Construction of a series of ompF-ompC chimeric genes by in vivo
RT homologous recombination in Escherichia coli and characterization of
RT the translational products.";
RN J. Bacteriol. 164:797-801(1985).
[7]
RN SEQUENCE OF 22-30.
RX STRAIN-K12 / EMG2;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded

```

```

RT In the genome of Escherichia coli K-12.";
RL Electrophoresis 18:1259-1313(1997).
[8]
RN SEQUENCE OF 22-26.
RX STRAIN-K12 / W3110;
RL MEDLINE=98291876; PubMed=9629924;
RA Molloy M.P., Herbert B.R., Walsh B.J., Tyler M.I., Traini M.,
RA Sanchez J.-C., Hochstrasser D.F., Williams K.L., Gooley A.A.;
RT "Extraction of membrane proteins by differential solubilization for
RT separation using two-dimensional gel electrophoresis.";
RL Electrophoresis 19:837-844(1998).
CC -!- FUNCTION: FORM PASSIVE DIFFUSION PORES WHICH ALLOW SMALL MOLECULAR
CC WEIGHT HYDROPHILIC MATERIALS ACROSS THE OUTER MEMBRANE.
CC -!- SUBUNIT: HOMOTRIMER.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -!- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; K00541; AAA24243.1; -.
DR EMBL; AE000310; AAC75275.1; -.
DR EMBL; D90850; BAA15998.1; -.
DR EMBL; U00008; AAL16412.1; -.
DR PIR; A20867; MWEPC.
DR PIR; B25029; B25029.
DR PIR; A18885; A18885.
DR HSP; O52503; IIIV.
DR ECO2DBASE; A035_5; 6TH EDITION.
DR EcoGene; EG10670; ompC.
DR InterPro; IPR003229; OMP_2.
DR InterPro; IPR001702; Porin_gram-ve.
DR Pfam; PF00267; Gram-ve_porins; 1.
DR PRINTS; PR00182; ECOLNEIPORIN.
DR PROSITE; PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Transmembrane; Porin; Phage recognition; Signal;
KW Complete proteome.
FT SIGNAL 1 21
FT CHAIN 22 367
FT CDS 22 367
SQ SEQUENCE 367 AA; 40368 MW; 6A49370CC8A1A225 CRC64;
-----
Query Match 58.1%; Score 1051; DB 1; Length 367;
Best Local Similarity 58.9%; Pred. No. 6.6e-69;
Matches 212; Conservative 45; Mismatches 69; Indels 34; Gaps 8;
-----
QY 1 AEIYNKDGKVDLYGKAVGLHYFSKNGNSYGGNGDMTYARLFGKGTQINSDLTGYGQ 60
DB ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
22 AEVYNKDGKLDLYGKVDGLHYFS-----DNKDVGDQDTYMRGLGFGETQVTDLTGYGQ 76
QY 61 WEYFNQGNNSGADAGTGNKTLAFAGLYADVGFGRNYGVYDYGALGYDMLPERFG 120
DB ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77 WEYIQGNSAEN---ENNSWTRVAFAGLKFQDVGSFGRNYGVYDVTWTDVLPFRFG 133
QY 121 DTAYSDDDFVGRVGVGATYRNPFGLVDGLNFAVQYLCKNERDT-----A 166
DB ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
134 DTYGSDNFMNQKNGFATYRNTDFGLVDGLNFAVQYQKNGNPGSGEGTSGVTNGRDA 193
QY 167 RRSNGDVGGSISYEYEGFIVGAYCAADRTNLQE-AOPLGNGKKAQWATLKKYDANNI 225
DB |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
194 LRQNGDVGGSITYDEYEGFICGAISSSKRTDAQNTAAIYINGCDRAEYTGGLKDYANNI 253
QY 226 YLAANYGETRNATPTTNKFTNTSGFANKTDQVLLVAQYQDFGLRPSIAYTKSKARDV-E 284
DB ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
254 YLAAQYTOTYNATRV-----GSLGWANKAQNFVAQYQDFGLRPSLAYLQSKCKNLGR 308
QY 285 GIGVDVLYNVEVGATYTFYFNKNMSTYVDYIINQIDSKNL-----CVGSDDTVAVGIVQF 340
DB ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
309 GYDDEDILKYVDVGATYTFYFNKNMSTYVDYKINLLD-DNQFTRDAGINTDNIVALGLVYQF 367

```

```
RESULT 10
PHOE_KLEPN STANDARD; PRT; 351 AA.
AC P30704;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Outer membrane pore protein E precursor.
GN PHOE.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87190417; PubMed=3032618;
RA van der Ley P., Bekkers A., van Meersbergen J., Tommassen J.;
RT "A comparative study on the phoE genes of three enterobacterial
RT species. Implications for structure-function relationships in a
RT pore-forming protein of the outer membrane.";
RL Eur. J. Biochem. 164:469-475 (1987).
CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN CELLS ARE
CC GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS
CC PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE,
CC PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
CC SOLUTES.
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE OMP/PHOE FAMILY OF PORINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M28295; AAA25121.1; -
CC HSSP; P02932; IPHO.
CC InterPro; IPR003229; OMP_2.
CC Pfam; PF00267; Gram-ve_porins; 1..
CC PRINTS; PR00182; ECOLNEIPORIN.
CC PROSITE; PS00576; GRAM_NEG_PORIN; 1.
CC Outer membrane; Transmembrane; Porin; Signal.
CC SIGNAL 1 21 BY SIMILARITY.
CC CHAIN 22 351 OUTER MEMBRANE PORE PROTEIN E.
CC SEQUENCE 351 AA; 38894 MW; 86A5286C12502EC7 CRC64;
Query Match 58.0%; Score 1050; DB 1; Length 351;
Best Local Similarity 58.6%; Pred. No. 7.4e-69;
Matches 201; Conservative 43; Mismatches 83; Indels 16; Gaps 5;
QY 1 AEYKNGKNDVLYGKAVGLHYESKNGENSGYCGNDMTYARLGFKEGTQINSDLTGCG 60
Db 22 AEYKNGKNDVLYGKAVGLHYESKNGENSGYCGNDMTYARLGFKEGTQINSDLTGCG 76
QY 61 WEYFQGNSEGAADQTKNTRLAFLAGLYADVGSFDYGRNYGVWYDALGYTDLMLPEFGG 120
Db 77 WESEFSGNKTESDSSQ--QKTRLAFLAGLYADVGSFDYGRNYGVWYDALGYTDLMLPEFGG 134
QY 121 D-TAYSDDFVGRGVGVATYRNSNFFGLVDGLNFAVOYLGKNERDPTARRNSGDVGSGSIS 179
Db 135 DSSAQTDNFWTKRASGLATYRNTDFGLVDGLTLQYQCKNEGRKAKNGDGVCTSL 194
QY 180 YEYEG--FGVYGAGADRTNLEQAPLNGKKAQWATGLKDYANNIYAAYGETRNA 237
Db 195 YDFGGDFVAAAYTSDDRTNDQNLRLAQAQSKAAWATGLKDYANNIYATMYSETRKM 254
QY 238 TPITNFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYTKSKADKVEGIDVDLVNFEV 297
Db 255 TPI-----SGGFANKAQNEAQAQYQDFGLRPSIATYTKSKADKVEGIDVDLVNFEV 308
QY 298 GATYYFNKNKNTYVDYIINQIDSNKLGVSDDTDVAVGIYQF 340
Db 309 GLTYYFNKNKNTYVDYIINQIDSNKLGVSDDTDVAVGIYQF 351
RESULT 11
PORI_BPPA2 STANDARD; PRT; 365 AA.
AC P07238;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane porin protein LC precursor.
GN LC.
OS Bacteriophage PA-2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC unclassified Siphoviridae.
OX NCBI_TaxID=10738;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86304457; PubMed=3017988;
RA Blasband A.J., Marcotte W.R. Jr., Schnaitman C.A.;
RT "Structure of the lc and nmpc outer membrane porin protein genes of
RT lambdaoid bacteriophage.";
RL J. Biol. Chem. 261:12723-12732 (1986).
CC -1- FUNCTION: PORINS ARE MAJOR PROTEINS FOUND IN THE OUTER MEMBRANES
CC OF GRAM-NEGATIVE BACTERIA WHERE THEY FORM CHANNELS FOR THE
CC NONSPECIFIC PERMEATION OF SMALL SOLUTES (MOLECULES WITH MW LOWER
CC THAN 4000-6000 DALTONS).
CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
CC (by similarity).
CC -1- MISCELLANEOUS: A PORIN GENE CAN ALSO BE FOUND IN THE GENOMES OF
CC CERTAIN LAMBDOID BACTERIOPHAGE, AND ITS PROTEIN IS EXPRESSED IN
CC THE LYCOGENIC STATE IN E. COLI. THE EXPRESSION OF OMP/PHOE
CC PROTEINS IS THEN REDUCED SUBSTANTIALLY.
CC -1- SIMILARITY: BELONGS TO THE OMP/PHOE FAMILY OF PORINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; J02580; AAA32301.1; -
CC PIR; D25647; MMBP2.
CC HSSP; O52503; IIIV.
CC InterPro; IPR003229; OMP_2.
CC InterPro; IPR001702; Porin-gram-ve.
CC Pfam; PF00267; Gram-ve_porins; 1.
CC PRINTS; PR00182; ECOLNEIPORIN.
CC PROSITE; PS00576; GRAM_NEG_PORIN; 1.
CC Outer membrane; Transmembrane; Porin; Signal.
CC SIGNAL 1 23
CC CHAIN 24 365 OUTER MEMBRANE PORIN PROTEIN LC.
CC CONFLICT 99 99 H -> R (IN REF. 1; AA SEQUENCE).
CC SEQUENCE 365 AA; 40290 MW; 0FBC0531F9C0205 CRC64;
Query Match 57.8%; Score 1046; DB 1; Length 365;
Best Local Similarity 59.5%; Pred. No. 1.5e-68;
Matches 210; Conservative 35; Mismatches 84; Indels 24; Gaps 7;
QY 1 AEYKNGKNDVLYGKAVGLHYESKNGENSGYCGNDMTYARLGFKEGTQINSDLTGCG 60
Db 24 AEYKNGKNDVLYGKAVGLHYESKNGENSGYCGNDMTYARLGFKEGTQINSDLTGCG 78
QY 61 WEYFQGNSEGAADQTKNTRLAFLAGLYADVGSFDYGRNYGVWYDALGYTDLMLPEFGG 120
Db 111 WEYFQGNSEGAADQTKNTRLAFLAGLYADVGSFDYGRNYGVWYDALGYTDLMLPEFGG 120
```





```
Db 22 AEVYKNDGKLDLYCKVDGLHYSFNSAK-----DGDQSYARLGFKGETQINDQITGYGQ 76
Qy 61 WEYNFQGNSEGADQATGNKTRAFAGLKYADVGSFSDYGRNYSVYDALGYTDMLPFEGG 120
Db 77 WEYNTQANNTESSKQSW--TRAFAGLKFADYGSFSDYGRNYSVYDALGYTDMLPFEGG 134
Qy 121 DT-AVSDDFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNE-----RDTRAR 167
Db 135 DSYTNADNFMTGRANGVATYRTDFGLVGLNFAVQYLGKNEGASNGQEGTNGRDR-VR 193
Qy 168 RNSGDSVGSISYEY-EGFGIVGAYGAADRTNLQEAQPLGNGKKAQWATGLKYDANNIY 226
Db 194 HENGDSGLSTYIDIGMGFSAGAAVTSDDRTNDQVNHHTAAGDGKADAWTAGLKYDANNIY 253
Qy 227 LAANGETERNATPITNKFTNTSGFANKQDVLVLAQYQDFGLRPSIATYKSKAKDVE-- 284
Db 254 LATMYSEIRNMTTFGD---SDYAVANKQNFETVTAQYQDFGLRPAVSELSKGRDLHNA 310
Qy 285 -----GIGDVLVNYFEVGATYFNKMSYVDYIINQIDSNKL-----CVGSDDTVA 333
Db 311 GGADNPAGVDDKDLVKYADIGATYFNKMSYVDYKINLDEDDSDFYAANGISTDDIVA 370
Qy 334 VGIVYQF 340
Db 371 LGLVYQF 377

RESULT 15
OMPF_SALTI
ID OMPF_SALTI STANDARD; PRT; 363 AA.
AC Q56113;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Outer membrane protein F precursor (Porin ompF) (Outer membrane
protein S3).
GN OMPF OR OMP53 OR STV1002.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_taxid=601;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=IMSS-1;
RA Fernandez-Mora M., Calva E.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Quail M., Rutherford K., Leather S., Moule S., O'Gaora P., Parry C.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
CC -!- FUNCTION: OMPF IS A PORIN THAT PASSIVE DIFFUSION PORES WHICH ALLOW
CC SMALL MOLECULAR WEIGHT HYDROPHILIC MATERIALS ACROSS THE OUTER
CC MEMBRANE. IT IS ALSO RECEPTOR FOR THE BACTERIOPHAGE T2 (BY
CC SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -!- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X89757; CAA61905.1; -.
DR EMBL; AL627268; CAD05399.1; -.
DR HSSP; P02931; IGFN.
DR InterPro; IPR003229; OMP_2.
DR InterPro; IPR001702; Porin_gram-ve.
DR Pfam; PF00267; Gram-ve_porins; 1.
DR PRINTS; PR00182; ECOLNEIPORIN.
DR PROSITE; PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Transmembrane; Porin; Phage recognition; Signal;
KW Complete proteome.
FT SIGNAL 1 22 BY SIMILARITY.
FT CHAIN 23 363 OUTER MEMBRANE PROTEIN F.
FT CONFLICT 307 307 D -> G (IN REF. 1).
SQ SEQUENCE 363 AA; 40106 MW; F5059B37EA516859 CRC64;

Query Match 57.0%; Score 1031.5; DB 1; Length 363;
Best Local Similarity 57.6%; Pred. No. 1.7e-67;
Matches 200; Conservative 54; Mismatches 80; Indels 13; Gaps 8;

Qy 1 AEIYNKDGKVDLYKAVGLHYFSKNGENSYGGNGDMTYARLGFKGETQINSIDLTGYGQ 60
Db 23 AEIYNKDGKLDLYGKAVGRHVWTT-TGDSK---NADQTYAQIGFGETQINTDLTGFGQ 78
Qy 61 WEYNFQGNSEGADQATGNKTRAFAGLKYADVGSFSDYGRNYSVYDALGYTDMLPFEGG 120
Db 79 WEYRTKADRAEG-EQONSNLVRLAFAGLKYAEVGSIDYGRNYSVYDMAPYFSG 137
Qy 121 DT---AYSDDFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRS-NGDGVGG 176
Db 138 ETWGGAYTDNYMTSRAGLLTYRNSDFFGLVDGLSFGIQYQCKNQDNHSINSQNGDGVGY 197
Qy 177 SISYEYEGFIVGAYGAADRTNLQEAQPLGNGKKAQWATGLKYDANNIYLAANYGETRN 236
Db 198 TMAYEFDGFGVTAAYSNSKRTNDQDRD-GNGDRAESWAVGAKYDANNVYLAAYTAETRN 256
Qy 237 ATPITNKFTNTSGFANKQDVLVLAQYQDFGLRPSIATYKSKAKDVEGI-GDVLVNYF 295
Db 257 MSIVENTVTDTVEMANKTONLEVVAQYQDFGLRPAISIVQSKQLNGADGSADLAKYI 316
Qy 296 EVGATYFNKMSYVDYIINQIDSNKLG--VGSDDTVAVGIVYQF 340
Db 317 QAGATYFNKMSYVDYRFLNLLDENDYSSSYVGTDDQAAVGITTYQF 363
```

Search completed: December 18, 2002, 16:18:23

Job time : 6.23408 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:13:04 ; Search time 20.0203 Seconds  
(without alignments)  
3509.535 Million cell updates/sec

Title: US-09-490-291-9  
Perfect score: 1809  
Sequence: 1 AEIYNKGNKVDLYGKAVGL.....NKLGVGSDDTVANGIVYQFA 341

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL.21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organalle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1789	98.9	362	16 Q8XDF1	Q8xdf1 escherichia
2	1390.5	76.9	351	2 Q93K98	Q93k98 enterobacte
3	1113	61.5	351	16 Q8X7N5	Q8x7n5 escherichia
4	1082	59.8	371	16 Q8ZGS6	Q8zgs6 yersinia pe
5	1061	58.7	363	2 Q93K99	Q93k99 enterobacte
6	1044.5	57.7	374	2 Q87754	Q87754 klebsiella
7	1036	57.3	377	2 Q85030	Q85030 escherichia
8	1035.5	57.2	372	2 Q93K39	Q93k39 klebsiella
9	1033	57.1	367	16 Q8XE41	Q8xe41 escherichia
10	1029.5	56.9	364	2 Q9K597	Q9k597 escherichia
11	1029.5	56.9	378	2 Q9K3E6	Q9k3e6 salmonella
12	1028.5	56.9	366	2 Q9RH85	Q9rh85 escherichia
13	1022.5	56.5	333	2 Q9RAW3	Q9raw3 klebsiella
14	1022.5	56.5	333	2 Q9K3E7	Q9k3e7 salmonella
15	1022.5	56.5	333	2 Q9RR59	Q9rr59 klebsiella
16	1021.5	56.5	333	2 Q9RR59	Q9rr59 klebsiella

17	1019.5	56.4	333	2 Q9S613	Q9s613 klebsiella
18	1019.5	56.4	333	2 Q8VDE7	Q8vde7 klebsiella
19	1019.5	56.4	333	2 Q8VPA5	Q8vpa5 klebsiella
20	1017.5	56.2	360	2 Q9F889	Q9f889 salmonella
21	1017	56.2	377	16 Q8ZPL4	Q8zpl4 salmonella
22	1007	55.7	375	2 Q9AGC9	Q9agc9 enterobacte
23	1005	55.6	375	2 Q9ALY0	Q9aly0 enterobacte
24	1000	55.3	375	2 Q9AGD1	Q9agd1 enterobacte
25	999	55.2	375	2 Q9AGC8	Q9agc8 enterobacte
26	998.5	55.2	366	16 Q8XAS0	Q8xas0 escherichia
27	998	55.2	375	2 Q93T24	Q93t24 enterobacte
28	995.5	55.0	376	2 Q9AGD0	Q9agd0 enterobacte
29	992	54.8	360	16 Q8ZG94	Q8z994 yersinia pe
30	984.5	54.4	398	16 Q8ZNS7	Q8zns7 salmonella
31	960	53.1	374	16 Q8ZGR1	Q8zgr1 yersinia pe
32	946	52.3	366	2 Q93SG6	Q93sg6 clostridium
33	938	51.9	342	2 Q8RLH4	Q8rlh4 salmonella
34	934.5	51.7	359	2 Q87753	Q87753 klebsiella
35	933	51.6	342	2 Q8RLH5	Q8rlh5 salmonella
36	933	51.6	342	2 Q8RLH3	Q8rlh3 salmonella
37	919	50.8	372	16 Q8ZPH7	Q8zph7 salmonella
38	815	45.1	315	2 Q9EXH8	Q9exh8 klebsiella
39	770	42.6	315	2 Q9ZC71	Q9zc71 yersinia pe
40	649.5	35.9	255	16 Q8X900	Q8x900 escherichia
41	539.5	29.8	224	16 Q8XB87	Q8xb87 escherichia
42	428.5	23.7	130	2 Q5Z641	Q5z641 rannella aq
43	423	23.4	191	16 Q8XB86	Q8xb86 escherichia
44	372.5	20.6	123	16 Q8X9Q2	Q8x9q2 escherichia
45	310	17.1	111	2 P94857	P94857 klebsiella

## ALIGNMENTS

## RESULT 1

Q8XDF1	PRELIMINARY;	PRT;	362 AA.
ID Q8XDF1			
AC Q8XDF1			
DT 01-MAR-2002 (Tremblrel. 20, Created)			
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)			
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)			
DE Outer membrane protein la (Ia,b,F).			
GN OMPF OR 21276 OR ECS1012.			
OS Escherichia coli O157:H7.			
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC Escherichia.			
OX NCBI_TaxID=83334;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;			
RX MEDLINE=21074935; PubMed=11206551;			
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,			
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,			
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,			
RA Grofbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,			
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,			
RA Welch R.A., Blattner F.R.;			
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.;"			
RL Nature 409:529-533(2001).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=O157:H7 / RMD 0509952;			
RX MEDLINE=21156331; PubMed=11258796;			
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,			
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,			
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,			
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;			
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli			
RT O157:H7 and genomic comparison with a laboratory strain K-12.;"			
RL DNA Res. 8:111-22(2001).			
DR EMBL; AF005283; AAC55414.1; "			
DR EMBL; AF005253; BAB34435.1; "			
DR InterPro; IPR003229; OMP_2.			

DR InterPro: IPR001702; Porin\_gram-ve.  
DR Pfam: PF00267; Gram-ve\_porins; 1.  
DR PRINTS; PR00182; ECOLNEIPORIN.  
DR PROSITE; PD000808; OMP\_2; 1.  
DR PROSITE; PS00576; GRAM\_NEG\_PORIN; 1.  
KW Complete proteome.  
SQ SEQUENCE 362 AA; 39361 MW; 9E6ACF4B9DAAB214 CRC64;

Query Match 98.9%; Score 1789; DB 16; Length 362;  
Best Local Similarity 99.4%; Pred. No. 2.5e-102;  
Matches 338; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AEIYNKDGKVDLYGKAVGLHYFSKNGSGNSYGGNDMTYARLGKGETQINSDLTGYGQ 60  
DB AEIYNKDGKVDLYGKAVGLHYFSKNGSGNSYGGNDMTYARLGKGETQINSDLTGYGQ 82  
QY 61 WEYFQGNNSGADAOQTKNKTFLAFAGLYADVGSDYGRNYYGVVYDALGYTDMLPFEGG 120  
DB WEYFQGNNSGADAOQTKNKTFLAFAGLYADVGSDYGRNYYGVVYDALGYTDMLPFEGG 142  
QY 121 DPAYSDDFFVGRVGVGATYRNSNFFGLVDGLNFAVOYLKGNKRDYARRSNGDVGGSISY 180  
DB DPAYSDDFFVGRVGVGATYRNSNFFGLVDGLNFAVOYLKGNKRDYARRSNGDVGGSISY 202  
QY 181 EYEGFGIYGAGAAADRTNLQEAQPLNGKKAQWATGLKYDANNIYLAANYGETRNATPI 240  
DB EYEGFGIYGAGAAADRTNLQEAQPLNGKKAQWATGLKYDANNIYLAANYGETRNATPI 262  
QY 241 TNKFTNTSGFANKTDVLLVAQYQDFGLRPSIATYTKSKAKDVEGIGDVLVNYFEVAT 300  
DB TNKFTNTSGFANKTDVLLVAQYQDFGLRPSIATYTKSKAKDVEGIGDVLVNYFEVAT 322  
QY 301 YFENKMTSYVDYIINQIDSDNKLGVGSDTVAVGIVYQF 340  
DB YFENKMTSYVDYIINQIDSDNKLGVGSDTVAVGIVYQF 362

RESULT 2  
Q93K98 PRELIMINARY; PRT; 351 AA.

ID Q93K98  
AC Ballesteros F., Domenech-Sanchez A., Martinez-Martinez L., Pascual A.,  
RA Cuneo M.C., Benedi V.J.;  
RA "Genetic characterization and role in ss-lactam resistance of the OmpC  
RT and ompF enterobacter cloacae porins";  
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ316540; CAC48383.1; -;  
DR InterPro; IPR003229; OMP\_2.  
DR InterPro; IPR001702; Porin\_gram-ve.  
DR Pfam; PF00267; Gram-ve\_porins; 1.  
DR PRODOM; PD000808; OMP\_2; 1.  
DR PROSITE; PS00576; GRAM\_NEG\_PORIN; UNKNOWN\_1.  
FT SIGNAL  
KW Porin; Signal.  
SQ SEQUENCE 351 AA; 38465 MW; 045617CB302F8968 CRC64;

Query Match 76.98%; Score 1390.5; DB 2; Length 351;  
Best Local Similarity 77.4%; Pred. No. 6.7e-78;  
Matches 263; Conservative 23; Mismatches 43; Indels 11; Gaps 2;

QY 1 AEIYNKDGKVDLYGKAVGLHYFSKNGSGNSYGGNDMTYARLGKGETQINSDLTGYGQ 60

DB AEIYNKDGKVDLYGKAVGLHYFSNDND-----GNDGDKTYARLGKGETKINDQLTGYGQ 77  
QY 61 WEYFQGNNSGADAOQTKNKTFLAFAGLYADVGSDYGRNYYGVVYDALGYTDMLPFEGG 120  
DB WEYFQGNNSGADAOQTKNKTFLAFAGLYADVGSDYGRNYYGVVYDALGYTDMLPFEGG 137  
QY 121 DPAYSDDFFVGRVGVGATYRNSNFFGLVDGLNFAVOYLKGNKRDYARRSNGDVGGSISY 180  
DB DPAYSDDFFVGRVGVGATYRNSNFFGLVDGLNFAVOYLKGNKRDYARRSNGDVGGSISY 197  
QY 181 EYEGFGIYGAGAAADRTNLQEAQPLNGKKAQWATGLKYDANNIYLAANYGETRNATPI 240  
DB EYEGFGIYGAGAAADRTNLQEAQPLNGKKAQWATGLKYDANNIYLAANYGETRNATPI 257  
QY 241 TNKFTNTSGFANKTDVLLVAQYQDFGLRPSIATYTKSKAKDVEGIGDVLVNYFEVAT 300  
DB TNKFTNTSGFANKTDVLLVAQYQDFGLRPSIATYTKSKAKDVEGIGDVLVNYFEVAT 311  
QY 301 YFENKMTSYVDYIINQIDSDNKLGVGSDTVAVGIVYQF 340  
DB YFENKMTSYVDYIINQIDSDNKLGVGSDTVAVGIVYQF 351

RESULT 3  
Q8X7N5 PRELIMINARY; PRT; 351 AA.  
ID Q8X7N5  
AC Q8X7N5;  
DT 01-MAR-2002 (Tremblrel. 20, Created)  
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Outer membrane pore protein E (E.Ic.NmpAB) (Outer membrane pore  
protein PhoE).  
GN PHOE OR 20302 OR EGS0268.  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
Rafaeli G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
Ratkovits E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
RL Nature 409:529-533(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / RMD 0509952;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
Han C.-G., Ohtsubo K., Nakayama K., Murata T., Tanaka M., Tobe T.,  
Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohemorrhagic Escherichia coli  
O157:H7 and genomic comparison with a laboratory strain K-12";  
RL DNA Res. 8:11-22(2001).  
DR EMBL; AE005202; ANG54566.1; -;  
DR EMBL; AP002551; BAB33691.1; -;  
DR InterPro; IPR003229; OMP\_2.  
DR InterPro; IPR001702; Porin\_gram-ve.  
DR Pfam; PF00267; Gram-ve\_porins; 1.  
DR PRINTS; PR00182; ECOLNEIPORIN.  
DR PRODOM; PD000808; OMP\_2; 1.  
DR PROSITE; PS00576; GRAM\_NEG\_PORIN; 1.  
KW Complete proteome.  
SQ SEQUENCE 351 AA; 38966 MW; 3E58F4F284B5D42E CRC64;

Query Match 61.58%; Score 1113; DB 16; Length 351;

Db	77	WEYFAANYAESQAK-DNKTRLAFAGLRYNLGSIDYGRNYGLVDIAAWTOMLPEFGN	135
Qy	121	DT-AYSDDFFVGRVGGVATYRNSNFFGLVDGLNFAVOYLCKN-----ERDTARRSN	170
Db	136	DSYTRDNFTGRTTGVATYRNTDFFGLVDGLKPSLQYQKNGAEGSTNNGRDTSKQ-N	194
Qy	171	GDVGGSISIEY-EGFGIVGAYGAADRTNLQEAQPLGNGKKAQWATGLKYDANNIYLAA	229
Db	195	GDGFLSSYEIGAGVSGAAYASSNRTLAQKNSTFGKDKADAWTGLKAYSDNGVYLA	254
Qy	230	NYGETRNPIT-----NKFTNTSGFANKTQDVLVLAQYQFQFGLRPSIAYTKSKAKDV	283
Db	255	NYAETRNMTPIGTFVAVNNVSTSYSGFANKTQIELVAQLDFGLKPSIAYTQSKGDI	314
Qy	284	EGIGDVLVNYFEVGYATYENKNNSTVDYIINQISDNKLGSGDDTVAAGVLYQF	340
Db	315	EGIGDVLVYDVGATYFENKNNSTVDYIKINQINDNKLNTONVVALGLVYQF	371
RESULT 5			
Q93K99	ID	Q93K99 PRELIMINARY; PRT; 363 AA.	
AC	Q93K99;		
DT	01-DEC-2001 (TremBrel. 19, Created)		
DT	01-DEC-2001 (TremBrel. 19, Last sequence update)		
DT	01-JUN-2002 (TremBrel. 21, Last annotation update)		
DE	OmpC porin precursor.		
GN	OmpC.		
OS	Enterobacter cloacae.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Enterobacter.		
OX	NCBI_TaxID=550;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ATCC3047;		
RA	Ballesteros F., Domenech-Sanchez A., Martinez-Martinez L., Pascual A.		
RA	Conejo M.C., Benedi V.J.		
RT	"Genetic characterization and role in ss-lactam resistance of the OmpC and ompF Enterobacter cloacae porins",		
RL	Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AJ316539; CAC48382.1; -		
DR	InterPro: IPR003229; OMP_2.		
DR	InterPro: IPR001702; Porin-gram-ve.		
DR	InterPro: IPR000504; RNA_rec_mot.		
DR	Pfam: PF00267; Gram-ve_porins; 1.		
DR	ProDom: PD00808; OMP_2; 1.		
DR	PROSITE; PS00576; GRAM_NEG_PORIN; UNKNOWN_1.		
DR	PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.		
KW	Porin; Signal.		
FT	SIGNAL		
SD	SEQUENCE 1 21 POTENTIAL.		
		363 AA; 39940 MW; 681370F63C3075B6 CRC64;	
Query Match			
Best Local Similarity 58.7%; Score 1061; DB 2; Length 363;			
Matches 215; Conservative 35; Mismatches 77; Indels 28; Gaps			
Qy	1	AEIYNKGNKVDLYGKAVGLHYSEKNGNSYGGCNDMTYARLGFKEQTINSDLTGYGQ	60
Db	22	AEIYNKGNKLDLYGKVDGLHYFSDDSQ-----DGDQTYRMLGFKEGTQVNDLTGYGQ	76
Qy	61	WEYFNQNNSEGADAQGNKTRLAFAGLRKYADYGSFDYGRNYGVVYDALGYDTMLPEFGG	120
Db	77	WEYIQGNNGEN---ENNSWTRVAFAGLKEFGDAGSPDYGRNYGVVYDTSWTDLPEFGG	133
Qy	121	DTAYSDDFFVGRVGGVATYRNSNFFGLVDGLNFAVOYLK-----NERDTARRSN	170
Db	134	DTYGSDFNMQRNGFATYRNSDFFGLVDGLNFAVOYQKNGSASGEDQTNNGRTELQN	193
Qy	171	GDVGGSISIEY-EGFGIVGAYGAADRTNLQEAQPLGNGKKAQWATGLKYDANNIYLAA	229
Db	194	GDVGGSITYNLGEFGIGTAVSSSKRTSQNDLTGNGDRAETVYGGKLYDANNIYLAA	253

```
Qy 230 NYGETRNTATPTNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIAYTSKAKDVE-GTGD 288
Db 254 QYTQTNATRVGN-----LGNWAKAQNEFVVAQYQDFGLRPSVAYLQSKGKDLNGYGF 308

Qy 289 VDLVNYFEVGATYYFNKNMSTYVDYIIQIDIS---DNKLGVSDDTVAAGVIYQF 340
Db 309 QDLKYVDVGATYYFNKNMSTYVDYKINLLDDKEFTRNAGISTDIDVALGLVYQF 363

RESULT 6
O87754
ID O87754 PRELIMINARY; PRT; 374 AA.
AC O87754;
DT 01-NOV-1998 (TremBrel. 08, Created)
DT 01-NOV-1998 (TremBrel. 08, Last sequence update)
DT 01-JUN-2002 (TremBrel. 21, Last annotation update)
DE OmpK37 porin precursor.
GN OMPK37.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_taxid=573;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SB8;
RX MEDLINE=99235749; PubMed=10217760;
RA Domenech-Sanchez A., Hernandez-Alles S., Martinez-Martinez L.,
RA Benedi V.J., Alberti S.;
RT "Identification and characterization of a new porin gene of Klebsiella
RT pneumoniae: Its role in beta-lactam antibiotic resistance.";
RL J. Bacteriol. 181:2726-2732(1999).
DR EMBL; AJ011502; CAA09666.1; -.
DR HSSP; Q48473; 10SM.
DR InterPro; IPR003229; OMP_2.
DR InterPro; IPR001702; Porin_gram-ve.
DR InterPro; IPR000408; Reg_chir_condens.
DR Pfam; PF00267; Gram-ve_porins; 1.
DR PRINTS; PD00182; ECOLNEIPORIN.
DR ProDom; PD000808; OMP_2; 1.
DR PROSITE; PS00626; RCCL_2; UNKNOWN_1.
KW Porin; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 374 OMPK37 PORIN.
SQ SEQUENCE 374 AA; 41612 MW; B66DF0D74A1A169B CRC64;

Query Match 57.7%; Score 1044.5; DB 2; Length 374;
Best Local Similarity 58.5%; Pred. No. 1.2e-56;
Matches 214; Conservative 37; Mismatches 76; Indels 39; Gaps 9;

Qy 1 AEIYNKGNKVDLYGKAVGLHYFSKNGENSGYGGNDMTYARLGPKEGTQINSDLTGYGQ 60
Db 22 AEIYNKGNKLDLYGKVDGLHYFSSDSKK-----DGDQTYLRFPGKEGTQINDILTYGQ 76

Qy 61 WEYNFQGNSEGAQAOTCNKTRLAFAGLYADVGSFDYGRNYGVYDLYGTYDMLPERGG 120
Db 77 WEYNQANNTETSSDAQW--TRLAFAGLYKVDYGSFDYGRNYGVYDLYGTYDMLPERGG 134

Qy 121 DT-AYSDFFYGRVGGVATYRNSNFFGLVDGLNFAVQYLGKNE-----RDFA 167
Db 135 DSYTNADNFMTRANGVATYRNTDFGLVNLNFAVQYQGNNEGASNGEGTNGCRD-VR 193

Qy 168 RNSGDVGGSISYEY-EGFGIVAGYGAADRNLQEAQPLGNGKKAQWATGLKYDANNIY 226
Db 194 HENGSGWGLSTYSYDF-GWISAAAAYTSSDRNTDQMTNARGDKAEAWTAGLKVD 253

Qy 222 ANNIYLAANYGETRNTATPTNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIAYTSKAK 281
Db 254 ANDIYLATMYSETRNTMTPGN-----DGVANKTQNFVTAQYQDFGLRPAISYLQSKG 308

Qy 282 DVEGIG----DVLVNYFEVGATYYFNKNMSTYVDYIIQIDISDNK-----GVGSDTVA 334
Db 309 DLYNNGRVADKDLVYMDVGATYYFNKNMSTYVDYKINLLDGNKDFYEDNGISTDINVAL 368
```

```
Qy 335 GIVYQF 340
Db 369 GLVYQF 374

RESULT 7
O85030
ID O85030 PRELIMINARY; PRT; 377 AA.
AC O85030;
DT 01-NOV-1998 (TremBrel. 08, Created)
DT 01-NOV-1998 (TremBrel. 08, Last sequence update)
DT 01-JUN-2002 (TremBrel. 21, Last annotation update)
DE Porin OmpN.
GN OMPN.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BE;
RX MEDLINE=98317278; PubMed=9642192;
RA Prilipov A., Phale P.S., Koebnik R., Widmer C., Rosenbusch J.P.;
RT "Identification and characterization of two quiescent porin genes,
RT nmpC and ompN, in Escherichia coli BE.";
RL J. Bacteriol. 180:3388-3392(1998).
DR EMBL; AF035618; AAC38644.1; -.
DR HSSP; P02931; IGFN.
DR InterPro; IPR003229; OMP_2.
DR InterPro; IPR001702; Porin_gram-ve.
DR InterPro; IPR000408; Reg_chir_condens.
DR Pfam; PF00267; Gram-ve_porins; 1.
DR PRINTS; PD00182; ECOLNEIPORIN.
DR ProDom; PD000808; OMP_2; 1.
DR PROSITE; PS00626; RCCL_2; UNKNOWN_1.
SQ SEQUENCE 377 AA; 41236 MW; 7F4D681A2BC10F8C CRC64;

Query Match 57.3%; Score 1036; DB 2; Length 377;
Best Local Similarity 57.2%; Pred. No. 4.1e-56;
Matches 210; Conservative 41; Mismatches 78; Indels 38; Gaps 9;

Qy 1 AEIYNKGNKVDLYGKAVGLHYFSKNGENSGYGGNDMTYARLGPKEGTQINSDLTGYGQ 60
Db 22 AEIYNKGNKLDLYGKVDGLHYFSDNSAK-----DGDQSYARLGPKEGTQINDQLTYGQ 76

Qy 61 WEYNFQGNSEGAQAOTCNKTRLAFAGLYADVGSFDYGRNYGVYDLYGTYDMLPERGG 120
Db 77 WEYNQANNTETSSKNQSW--TRLAFAGLYKVDYGSFDYGRNYGVYDIEGWTDLPERGG 134

Qy 121 DT-AYSDFFYGRVGGVATYRNSNFFGLVDGLNFAVQYLGKNE-----RDFA 167
Db 135 DSYTNADNFMTRANGVATYRNTDFGLVNLNFAVQYQGNNEGASNGEGTNGCRD-VR 193

Qy 168 RNSGDVGGSISYEY-EGFGIVAGYGAADRNLQEAQPLGNGKKAQWATGLKYDANNIY 226
Db 194 HENGSGWGLSTYDLYGMSGFASAGAAVTSSTDRNDQVNHHTAAGDKADAWTAGLKVDANNIY 253

Qy 227 LAANYGETRNTATPTNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIAYTSKAKDVE-- 284
Db 254 LATMYSETRNTMTPGDD---SDYAVANKTQNFVTAQYQDFGLRPAVSLMSKGRDLHAA 310

Qy 285 -----GIGDVLVNYFEVGATYYFNKNMSTYVDYIIQIDISDNK-----GVGSDTVA 333
Db 311 GGADNPAGVDKDLVYADVGATYYFNKNMSTYVDYKINLLDEDDSFYTANGISTDIDVA 370

Qy 334 GVIVYQF 340
Db 371 LGLVYQF 377

RESULT 8
Q93K39
ID Q93K39 PRELIMINARY; PRT; 372 AA.
```

```
AC Q93K39;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE OmpK36 protein precursor.
GN OmpK36.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=103624;
RA Crowley B., Benedi V.J., Domenech-Sanchez A.;
RT "Porin deficiency and SHV-2 expression results in increased resistance
RT to cephalosporins and carbapenems in Klebsiella pneumoniae.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ344089; CAC50885.1; -
DR InterPro: IPR003229; Omp.2.
DR Pfam: PF00267; Gram-ve_porins; 1.
DR ProDom: PD000808; Omp_2; 1.
DR PROSITE: PS00576; GRAM_NEG_PORIN; UNKNOWN_1.
KW SIGNAL.
FT CHAIN
FT SIGNAL
SQ SEQUENCE 372 AA; 40753 MW; 94B84DA54AF97736 CRC64;
Query Match 57.2%; Score 1035.5; DB 2; Length 372;
Best Local Similarity 58.4%; Pred. No. 4.3e-56;
Matches 209; Conservative 38; Mismatches 86; Indels 25; Gaps 6;
QY 1 AEYTKDGNKVDLYGKAVGLHYFSKNGENSYGNGDMTYARLFGKGTQINSLDTGYGQ 60
DB 22 AEYTKDGNKLDLYGKIDGLHYFS-----DDKSVDDGQTYMRVGKGTQINDLTGYGQ 76
QY 61 WEYNFQGNNSGADQGTNTRLAFAGLYADVGSFDYGRNYYGVVYDALGYTDMLPFEGG 120
DB 77 WEYNQANNTSSSDQAW--TRLAFAGLKFGDAGSFDYGRNYYGVVDTWTDVLPFEGG 134
QY 121 DTAYSDDFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDT-----ARR 168
DB 135 DTYSDFNQLSRANGVATYRNSDFGLVDGLNFAVQYLGKNGSISGEGSTPNNGRGALK 194
QY 169 SNGDVGGSISYE-YEGFIVGAYGAADRNLQEAQPLNGKKAQEWATGLKYDANNIYL 227
DB 195 QNGDGFGLTSLTYDYGISAGFYACNSKRNGDQNRDLKGRGDNAETTYTGLKYDANNIYL 254
QY 228 AANYGETRNATPIT--NKFTNTSGFANKTODVLLVAQYQDFGLRPSIATYTKSAKDVEG 285
DB 255 ATQYQTYNATRV-----GSLGWANKAQNFVAQYQDFGLRPSVAYLQSKGKDIEG 314
QY 286 IGDVLDVNYFEVGATYFENKMSYVDYIINQIDSDN---KLGVGSDDTVAAGVIYQF 340
DB 315 YGDQDLKYVDVGATYFENKMSYVDYKINLLDENDFTRSGAGISTDDVVALGLVYQF 372
RESULT 9
Q8XE41 ID Q8XE41 PRELIMINARY; PRT; 367 AA.
AC Q8XE41;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Outer membrane protein 1b (Ib,c) (Outer membrane protein C OmpC).
GN OmpC OR Z3473 OR ECS3104.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
```

```
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Gostfal G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Probeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shigenaga H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL: AE005453; AAG57350.1; -
DR EMBL: AP002560; BAB36527.1; -
DR InterPro: IPR003229; Omp.2.
DR InterPro: IPR001702; Porin_gram-ve.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF00267; Gram-ve_porins; 1.
DR PRINTS: PR00182; ECOLNEIPORIN.
DR ProDom: PD000808; Omp_2; 1.
DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
DR PROSITE: PS00030; RRM_RNP_1; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 367 AA; 40508 MW; 028A0146D462CAA4 CRC64;
Query Match 57.1%; Score 1033; DB 16; Length 367;
Best Local Similarity 58.7%; Pred. No. 6e-56;
Matches 212; Conservative 43; Mismatches 70; Indels 36; Gaps 9;
QY 1 AEYTKDGNKVDLYGKAVGLHYFSKNGENSYGNGDMTYARLFGKGTQINSLDTGYGQ 60
DB 22 AEYTKDGNKLDLYGKIDGLHYFS-----DDKSVDDGQTYMRVGKGTQINDLTGYGQ 76
QY 61 WEYNFQGNNSGADQGTNTRLAFAGLYADVGSFDYGRNYYGVVYDALGYTDMLPFEGG 120
DB 77 WEYQIQNSAEN---ENNSWTRVAFAGLKFDQVGSFDYGRNYYGVVDTWTDVLPFEGG 133
QY 121 DTAYSDDFVGRVGVATYRNSNFFGLVDGLNFAVQYLGK-----NERDTARRSN 170
DB 134 DTYSDFNQLSRANGVATYRNTDFGLVDGLNFAVQYLGKNGSVSGEGMTNNGREALQN 193
QY 171 GDVGGSISYEYEGFIVGAYGAADRNLQEAQPL--GNGKKAQEWATGLKYDANNIYLA 228
DB 194 GDVGGSISYDYEGFIVGAYGAADRNLQEAQPL--PLYIGNGDRAETTYTGLKYDANNIYLA 252
QY 229 ANYGETRNATPITNKFTNTSGFANKTODVLLVAQYQDFGLRPSIATYTKSAKDVEGIG- 287
DB 253 AQYQTYNATRV-----GSLGWANKAQNFVAQYQDFGLRPSLAYLQSKNGKLVING 307
QY 288 ----DVLVNYFEVGATYFENKMSYVDYIINQIDSDNKL----GVGSDDTVAAGVIYQ 339
DB 308 RNYDDEDILKYVDVGATYFENKMSYVDYKINLLD--DNQFTRDAGINTDNIVALGLVYQ 366
QY 340 F 340
DB 367 F 367
RESULT 10
Q9K597 ID Q9K597 PRELIMINARY; PRT; 364 AA.
AC Q9K597;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
```

ID	Q9RH85	PRELIMINARY;	PRT;	366 AA.
DC	Q9RH85			
DT	01-MAY-2000	(TREMBLrel. 13, Created)		
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)		
DT	01-JUN-2000	(TREMBLrel. 21, Last annotation update)		
DE	Outer membrane protein OmpC.			
GN	OMPc.			
OS	Escherichia coli O157:H7.			
OC	Bacteria; Proteobacteria; gamma subdlvision; Enterobacteriaceae;			
OC	Escherichia.			
OX	NCBI_TaxId=83334;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=O157:H7;			
RA	Yu S.L., Syu W.J.;			

\*Altered outer membrane protein OmpC in hemorrhagic Escherichia coli O157:H7\*;  
Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE (BY SIMILARITY).  
DR EMBL: AF057355; AAF21761.1; -.  
DR HSSP: O52503; 1IIV.  
DR InterPro: IPR003229; OMP\_2.  
DR InterPro: IPR001702; Porin\_gram-ve.  
DR Pfam: PF00267; Gram-ve\_porins; 1.  
DR PRINTS: PR00182; ECOLNEIPORIN.  
DR PROSITE: PS00576; GRAM\_NEG\_PORIN; 1.  
DR PRODOM: PD000808; OMP\_2; 1.  
DR PROSITE: PS00576; GRAM\_NEG\_PORIN; 1.  
KW Outer membrane; Porin; Signal; Transmembrane.  
SQ SEQUENCE 366 AA; 40499 MW; 6A4EAD1652565C00 CRC64;

Query Match 56.9%; Score 1028.5; DB 2; Length 366;  
Best Local Similarity 58.6%; Pred. No. 1.1e-55;  
Matches 211; Conservative 43; Mismatches 71; Indels 35; Gaps 9;

QY 1 AEIYNKDGKVDLYGRAVGLHYESKNGENSYGGNGDMTYARLGFKGETQINSDLTGYGQ 60  
DB 22 AEIYNKDGKLDLYGKVDGLHYES-----DDKSVDSGDTYMRLLGFKGETQVTDQLTGYGQ 76  
QY 61 WEYNFQGNSEGADAQTGNKTRLAFAAGLKYADYGSFSGYGRNYGVVVDALGYTDMLPFEGG 120  
DB 77 WEYIOGNSAEN---ENNSWTRVAFAGLKFQDVGSEFDYGRNYGVVVDVTSWTDVLPFEGG 133  
QY 121 DTAYSDDFFVGRVGGVATYNSNFFGLVDGLNFAVQYLGK-----NERDTARRSNG 171  
DB 134 DTYGSDNFQMQRNGFATYRTDFGLVDGLNFAVQYQCKNGSVSEGMTNNGREALRQNG 193  
QY 172 DVGGSISYEYEGFIVGAYGAADRNLQEAQPL--GNGKKAQWATGLKYDANNIYLA 229  
DB 194 DVGGSITDYEGFICAAVSSSKRTDDQNS-PLYIGNGDRAETYTGLKYDANNIYLA 252  
QY 230 NYGETRATPITNKTSGFANKTODVLLVAQYQDFGLRPSIATYTKSKAKDVEGIG-- 287  
DB 253 QYQTQTNATRV-----GSLGWAKNAQNFVAQYQDFGLRPSLAYLQSKGNLGVNGR 307  
QY 288 ---DVDLVNYFEVGATYFNKNKSTVDYIINOISDNKL-----GVGSDTAVGVIYQF 340  
DB 308 NYDDEDILKYVDVGATYFNKNKSTVDYIYNKLLD-DNQFTRDAGINTDNVALGLFYQF 366

RESULT 13  
Q9RAW3  
ID Q9RAW3 PRELIMINARY; PRT; 333 AA.  
AC Q9RAW3  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Phosphate porin (Fragment).  
GN PHOE.  
OS Klebsiella pneumoniae (subsp. ozaenae).  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Klebsiella.  
OX NCBI\_Taxid=574;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1436;  
RA Carter J.S., Bowden F.J., Bastian I., Myers G.M., Sriprakash K.S., Kemp D.J.;  
RA "Phylogenetic analysis of Calymmatobacterium granulomatis.";  
RT Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
RL -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE (BY SIMILARITY).  
CC EMBL: AF009229; AAD21517.1; -.  
DR HSSP: P02932; 1PHO.  
DR InterPro: IPR003229; OMP\_2.  
DR InterPro: IPR001702; Porin\_gram-ve.

DR PFAM: PF00267; Gram-ve\_porins; 1.  
DR PRINTS: PR00182; ECOLNEIPORIN.  
DR PRODOM: PD000808; OMP\_2; 1.  
DR PROSITE: PS00576; GRAM\_NEG\_PORIN; 1.  
KW Outer membrane; Porin; Signal; Transmembrane.  
SQ SEQUENCE 333 AA; 36839 MW; 68AF347207770ABA CRC64;

Query Match 56.5%; Score 1022.5; DB 2; Length 333;  
Best Local Similarity 58.9%; Pred. No. 2.4e-55;  
Matches 196; Conservative 41; Mismatches 79; Indels 17; Gaps 5;

QY 1 AEIYNKDGKVDLYGRAVGLHYESKNGENSYGGNGDMTYARLGFKGETQINSDLTGYGQ 60  
DB 15 AEIYNKANKLDYGIKAMHYFSDYDSK-----DGDQTYVRFGIKGETQINDDLTGYGR 69  
QY 61 WEYNFQGNSEGADAQTGNKTRLAFAAGLKYADYGSFSGYGRNYGVVVDALGYTDMLPFEGG 120  
DB 70 WESEFSGNKTESDSSQ---KTRLAFAAGVGLKNGYSGFDYGRNLGALYDVEAWTDMFPFEGG 126  
QY 121 D-TAYSDDFFVGRVGGVATYNSNFFGLVDGLNFAVQYLGKNERDTRARRSNGDVGSGSIS 179  
DB 127 DSSAQTDNFTKRAASGLATYRTDFGLVDGLDLTLLQYQCKNEGKAKQNGDVGVTSL 186  
QY 180 YEYEG--FGIVGAYGAADRNLQEAQPLGNGKKAQWATGLKYDANNIYLAANYGETRNA 237  
DB 187 YDEGSGDFAYSAAYTSSDRTDQNLARGOGSKAEAWATGLKYDANNIYLATYSETRKM 246  
QY 238 TPITNKTNTSGFANKTODVLLVAQYQDFGLRPSIATYTKSKAKDVEGIGDVLVNYFEV 297  
DB 247 TPI-----SGGFANKNAQNFVAQYQDFGLRPSLGYVLGSKKDIEGVGSEDLVNI 300  
QY 298 GATYFNKNKSTVDYIINOISDNKLGVGSD 330  
DB 301 GLTYFNKNMNAFVDYKINQLKSDNKLGINDD 333

RESULT 14  
Q9RAW2  
ID Q9RAW2 PRELIMINARY; PRT; 333 AA.  
AC Q9RAW2  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Phosphate porin (Fragment).  
GN PHOE.  
OS Klebsiella pneumoniae subsp. rhinoscleromatis.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Klebsiella.  
OX NCBI\_Taxid=39831;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCTC5046;  
RA Carter J.S., Bowden F.J., Bastian I., Myers G.M., Sriprakash K.S., Kemp D.J.;  
RA "Phylogenetic evidence for reclassification of Calymmatobacterium granulomatis as Klebsiella granulomatis comb. nov.";  
RT Int. J. Syst. Bacteriol. 49:1695-1700(1999).  
RL -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE (BY SIMILARITY).  
CC EMBL: AF009230; AAD21518.1; -.  
DR HSSP: P02932; 1PHO.  
DR InterPro: IPR003229; OMP\_2.  
DR InterPro: IPR001702; Porin\_gram-ve.  
DR PFAM: PF00267; Gram-ve\_porins; 1.  
DR PRINTS: PR00182; ECOLNEIPORIN.  
DR PRODOM: PD000808; OMP\_2; 1.  
DR PROSITE: PS00576; GRAM\_NEG\_PORIN; 1.  
KW Outer membrane; Porin; Signal; Transmembrane.  
SQ SEQUENCE 333 AA; 36839 MW; 68AF347207770ABA CRC64;

FT NON\_TER 333 333  
SQ SEQUENCE 333 AA; 36839 MW; 68AF347207770ABA CRC64;  
Query Match 56.5%; Score 1022.5; DB 2; Length 333;  
Best Local Similarity 58.9%; Pred. No. 2.4e-55;  
Matches 196; Conservative 41; Mismatches 79; Indels 17; Gaps 5;  
QY 1 ABIYKDKGNKVDLYGKAVGLHYFSGKNGENSYGGNDMTYARLFGKGETQINSDLTGYGQ 60  
DB 15 ABEYKDKGNKVDLYGKAVGLHYFSGKNGENSYGGNDMTYARLFGKGETQINSDLTGYGQ 69  
QY 61 WEYNQGNNSGADAGTGNKTRLAFLAGLYADYGVGDFYGRNYYGVVYDALGYDMLPFGG 120  
DB 70 WESESGNKTESDSSQ---KTLAFAGVLYKNGYSGDFYGRNYYGVVYDALGYDMLPFGG 126  
QY 121 D-TAYSDDFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTPARRSDYGVGGSIS 179  
DB 127 DSAQTDNFWTRASGLATYRTDFGLVDGLDLILQYQKNEGKQNGDGVGTSLS 186  
QY 180 YEYEG--FGIVGAYGAADRNLQEAQPLGNGKKAQWATGLKYDANNIYLAANYGETRNA 237  
DB 187 YDFGSGDFAVSAAYTSSDRNTDQNLARGQSKAEAWATGLKYDANNIYLATMYSETRKM 246  
QY 238 TPTTNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYKSKAKDVEGIDVLYNYFEV 297  
DB 247 TPI-----SGGFANKAQNFVAQYQDFGLRPSIATYKSKAKDVEGIDVLYNYFEV 300  
QY 298 GATYYFNKNMSTYVDYIINQIDSNKLGVSDD 330  
DB 301 GLTYFFNKNMNAFVYKINQLKSDNKLINDDD 333  
RESULT 15  
Q9K3E7  
ID Q9K3E7 PRELIMINARY; PRT; 363 AA.  
AC Q9K3E7;  
DT 01-OCT-2000 (TremBLrel. 15, Created)  
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)  
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)  
DE Outer membrane protein Fx precursor.  
GN OMPFX.  
OS Salmonella enterica subsp. enterica serovar Minnesota.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=70803;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SF1111;  
RA Zimmermann H., Wassenaar T.M., Laubenheimer-Preusse H., Petry F.,  
RA Loos M.;  
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Y15843; CAB96614.1; -.  
DR HSP; Y02931; IGPN.  
DR InterPro; IPR003229; OMP\_2.  
DR InterPro; IPR001702; Porin\_gram-ve.  
DR Pfam; PF00267; Gram-ve\_porins; 1.  
DR PRINTS; PR00182; ECOLNEIPORIN.  
DR ProDom; PD000808; OMP\_2; 1.  
DR PROSITE; PS00576; GRAM\_NEG\_PORIN; UNKNOWN\_1.  
KW Signal.  
FT SIGNAL. 1 22 POTENTIAL.  
FT CHAIN 23 363 OUTER MEMBRANE PROTEIN FX.  
SQ SEQUENCE 363 AA; 40090 MW; F5058DEDEA516859 CRC64;  
Query Match 56.5%; Score 1022.5; DB 2; Length 363;  
Best Local Similarity 57.3%; Pred. No. 2.6e-55;  
Matches 199; Conservative 54; Mismatches 81; Indels 13; Gaps 8;  
QY 1 ABIYKDKGNKVDLYGKAVGLHYFSGKNGENSYGGNDMTYARLFGKGETQINSDLTGYGQ 60  
DB 23 ABEYKDKGNKVDLYGKAVGRHWVTT-TGDSK---NADQTYAQIGKGETQINTDLTGFCQ 78  
QY 61 WEYNQGNNSGADAGTGNKTRLAFLAGLYADYGVGDFYGRNYYGVVYDALGYDMLPFGG 120

DB 79 WEYRTKADRAEG-EQONSNLVRLAFAGLYAEVGSIDYGRNYYGVVDSYTDMAFYEG 137  
QY 121 DT---AYSDDFFVGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERDTPARRS-NGDGVGG 176  
DB 138 ETWGGAYTDNTYTSRAGGLTYRNSDFFGLVDGLSFGIOYQGNQDNHHSINSQSGDVG 197  
QY 177 SISYEYEGFIVGAYGAADRNLQEAQPLGNGKKAQWATGLKYDANNIYLAANYGETRN 236  
DB 198 TWAYEFDGFGVTAAYSNSKRTNDQDRD-GNGDRAESWAGAKYDANNVYLAAYVATRN 256  
QY 237 APTPNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYKSKAKDVEGI-GDVLVNYF 295  
DB 257 MSIVENTVTDTVMANKTQNLVVAQYQDFGLRPSIATYKSKAKDVEGI-GDVLVNYF 316  
QY 296 EVGATYYFNKNMSTYVDYIINQIDSNKLG--VGSDDTVAVGIVYQF 340  
DB 317 QAGATYYFNKNMNVVYRFRNLLDENYSSYVGTVDVQAAVGITYQF 363

Search completed: December 18, 2002, 16:21:01  
Job time : 22.0203 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:05:10 : Search time 4.94091 Seconds  
(without alignments)  
1995.696 Million cell updates/sec

Title: US-09-490-291-11

Perfect score: 377

Sequence: 1 MRGSHHHHGHGSMASGDLKN.....AKIGDLNNTSGIRRPAAKLN 74

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_101002:\*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	377	100.0	74	22	AA82613
2	279.5	74.1	76	21	AA810562
3	243.5	64.6	76	21	AA810561
4	221	58.6	84	19	AAW43025
5	213	56.5	84	19	AAW43026
6	205	54.4	84	19	AAW43027
7	199	52.8	42	21	AA810558
8	197	52.3	84	19	AAW43028
9	181	48.0	84	19	AAW43024
10	159	42.2	42	21	AA810557

11	124	32.9	42	19	AAW43029
12	121	32.1	43	21	AA808364
13	121	32.1	43	21	AA808383
14	119.5	31.7	103	22	AA870797
15	118.5	31.4	93	22	AA870796
16	116	30.8	94	22	AA86459
17	115	30.5	43	21	AA808344
18	114.5	30.4	43	21	AA808375
19	113	30.0	43	21	AA808362
20	113	30.0	43	21	AA808366
21	111.5	29.6	43	21	AA808386
22	108	28.6	94	22	AA86460
23	104	27.6	43	21	AA808378
24	103	27.3	113	23	AA883407
25	101.5	26.9	286	21	AA556997
26	101.5	26.9	286	21	AA559578
27	101	26.8	117	23	AA883405
28	97	25.7	291	22	AA882608
29	97	25.7	681	22	AA882609
30	96.5	25.6	491	18	AAW13373
31	96	25.5	175	21	AA444383
32	95.5	25.3	128	22	AA866825
33	95.5	25.3	128	22	AA866828
34	95	25.2	373	14	AA839486
35	94	24.9	168	21	AAV44384
36	94	24.9	169	21	AAV44380
37	94	24.9	175	21	AAV44382
38	92	24.4	145	20	AAW94266
39	92	24.4	159	19	AAW31552
40	92	24.4	493	12	AA813992
41	91.5	24.3	168	21	AAV44385
42	91.5	24.3	169	21	AAV44381
43	91	24.1	18	21	AA810560
44	91	24.1	379	10	AA890508
45	90.5	24.0	380	10	AA890509

#### ALIGNMENTS

RESULT 1

AA82613

ID AA82613 standard; Protein; 74 AA.

AC AA82613;

XX

DT 02-OCT-2001 (first entry)

DE Recognin B1 protein.

DE Recognin B1; structural protein; purification; fibre; spinning.

OS Synthetic.

XX

XX

PN W0200153333-1.

PD 26-JUL-2001.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX



PT Synthetic block copolymer with an antibiotic compound, useful for  
 PT dressing abrasion, burn or non-puncture wound, comprises two  
 PT alpha-helical protein blocks, random-coil protein block and linker  
 XX proteins  
 PS  
 PS Example 1; Column 33-34; 53pp; English.  
 XX  
 CC This invention describes a novel synthetic block copolymer XYZ comprising  
 CC two alpha-helical protein blocks X and Z, a water soluble, random-coil  
 CC protein block Y which links X and Z, and linker proteins for linking the  
 CC alpha-helical protein blocks to the random-coil protein block. The  
 CC products of the invention have vulnary activity. A copolymer produced  
 CC from a gene that encodes a derivative of AC 1.0A copolymer was suspended  
 CC in water to form a gel, which was then used to treat a wound. The cell  
 CC binding domain was the integrin ArgGlyAspSer sequence which is known to  
 CC bind to gp130/IL6 proteins expressed on fibroblast cells necessary for  
 CC matrix formation for the regeneration of skin. Fibroblasts become  
 CC entrapped within the gel and thus remain at the site of the wound,  
 CC serving as a scaffold for the regeneration of tissue surrounding the  
 CC wound. A synthetic block copolymer together with an antibiotic compound  
 CC is useful for wound dressing. The copolymer is useful for dressing  
 CC abrasion, burn or non-puncture wound. Unlike conventional polymers, the  
 CC polymers of equal size are produced biologically from a single template.  
 CC The intermolecular binding of alpha-helical blocks are monodisperse which  
 CC gives a uniform pore size depending on the length of random-coil block.  
 CC The new copolymers have low molecular weight hence have decreased  
 CC viscosity solutions or gels under suitable conditions. This sequence  
 CC represents a copolymer block protein L2-A which is used to illustrate the  
 CC method of the invention.  
 XX  
 SQ Sequence 76 AA;

Query Match 64.6%; Score 243.5; DB 21; Length 76;  
 Best Local Similarity 75.0%; Pred. No. 1.4e-16;  
 Matches 48; Conservative 10; Mismatches 1; Indels 5; Gaps 1;  
 QY 1 MRGSHHHHHHGS-----MASGDLKNKVAQLKRRVSLKDKAAELKQVSRLENEIEDLKA 55  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 Db 1 MRGSHHHHHHGSDDDDKASGDLNEVAQLEREVRSLEDEAAELEQKVRSLKNEIEDLKA 60  
 QY 56 KIGD 59  
 :|||  
 Db 61 EIGD 64

RESULT 4  
 AAW43025  
 ID AAW43025 standard; Peptide; 84 AA.  
 XX  
 AC AAW43025;  
 XX  
 DT 08-MAY-1998 (first entry)  
 XX  
 DE Artificial recognition sequence 2.  
 XX  
 KW Recognition sequence; self-assembling; protein-based structural material;  
 KW spontaneous heterodimerisation; coiled-coil alpha helix;  
 KW selectively permeable membrane; coated fabric.  
 XX  
 OS Synthetic.  
 XX  
 PN US5712366-A.  
 XX  
 PD 27-JAN-1998.  
 XX  
 XX 25-MAY-1995; 95US-0452592.  
 XX  
 PF 25-MAY-1993; 93US-0068948.  
 XX  
 PR (USSA ) US SEC OF ARMY.  
 PA Kaplan DL, McGrath KP;  
 XX  
 PI WPI; 1998-120029/11.  
 XX  
 PT Production of peptide hetero:dimer(s) - in the fabrication of  
 PT self-assembling protein-based structural material(s)

DR WPI; 1998-120029/11.  
 XX  
 PT Production of peptide hetero:dimer(s) - in the fabrication of  
 PT self-assembling protein-based structural material(s)  
 XX  
 PS Disclosure; Fig 4B; 17pp; English.  
 XX  
 CC AAW43024-28 represent recognition sequences obtained from the library of  
 CC recognition sequences coded for by AAV04943. This DNA sequence is  
 CC obtained by reverse translation. The codons were chosen to maximise  
 CC expression in Escherichia coli, and to introduce useful restriction for  
 CC subsequent genetic manipulations. The DNA sequence use a "mixed site"  
 CC approach at the first base of the codons for amino acids positions 5 and  
 CC 7 of the heptad (see below). The formation of a self-assembling  
 CC structural material is mediated by the artificial recognition sequences  
 CC (e.g. present sequence). A method of producing a self-assembling  
 CC protein-based structural material comprises mixing a multiplicity of  
 CC 2 different artificial peptide sequences, designed so as to  
 CC spontaneously heterodimerise. Both artificial peptides have a coiled-coil  
 CC alpha helical secondary structure having at least 2 heptad units. The  
 CC artificial peptides spontaneously heterodimerise into a self-assembled  
 CC protein-based structural material. The method is used to produce  
 CC nanoscale structural materials designed for specific functions e.g.  
 CC membranes, fibres, absorbants, reactive materials, etc. A specific  
 CC application is in the fabrication of selectively permeable membranes for  
 CC protective gloves and coated fabrics for fuel handlers and other  
 CC hazardous materials. Materials for medical implants, wound healants and  
 CC other medical treatments would also be feasible using this technology.  
 XX  
 SQ Sequence 84 AA;

Query Match 58.6%; Score 221; DB 19; Length 84;  
 Best Local Similarity 97.9%; Pred. No. 3.9e-16;  
 Matches 46; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 16 GDLKNKVAQLKRRVSLKDKAAELKQVSRLENEIEDLKAIGDLNN 62  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 Db 2 GDLKNKVAQLKRRVSLKDKAAELKQVSRLENEIEDLKAIGDLKN 48

RESULT 5  
 AAW43026  
 ID AAW43026 standard; Peptide; 84 AA.  
 XX  
 AC AAW43026;  
 XX  
 DT 08-MAY-1998 (first entry)  
 XX  
 DE Artificial recognition sequence 3.  
 XX  
 KW Recognition sequence; self-assembling; protein-based structural material;  
 KW spontaneous heterodimerisation; coiled-coil alpha helix;  
 KW selectively permeable membrane; coated fabric.  
 XX  
 OS Synthetic.  
 XX  
 PN US5712366-A.  
 XX  
 PD 27-JAN-1998.  
 XX  
 XX 25-MAY-1995; 95US-0452592.  
 XX  
 PF 25-MAY-1993; 93US-0068948.  
 XX  
 PR (USSA ) US SEC OF ARMY.  
 PA Kaplan DL, McGrath KP;  
 XX  
 PI WPI; 1998-120029/11.  
 XX  
 PT Production of peptide hetero:dimer(s) - in the fabrication of  
 PT self-assembling protein-based structural material(s)







PT Production of peptide heterodimer(s) - in the fabrication of  
PT self-assembling protein-based structural material(s)

XX Claim 2; Columns 13-14; 17pp; English.

XX The present sequence is used to obtain an entire library of recognition  
CC sequences. The DNA sequence encoding the present sequence is obtained by  
CC reverse translation. The codons were chosen to maximise expression in  
CC Escherichia coli, and to introduce useful restriction for subsequent  
CC genetic manipulations. The DNA sequence use a "mixed site" approach at  
CC the first base of the codons for amino acids positions 5 and 7 of the  
CC heptad (see below). The formation of a self-assembling structural  
CC material is mediated by the artificial recognition sequences obtained  
CC from the present sequence. A method of producing a self-assembling  
CC protein-based structural material comprises mixing a multiplicity of  
CC 2 different artificial peptide sequences, designed so as to spontaneously  
CC heterodimerise. Both artificial peptides have a coiled-coil alpha helical  
CC secondary structure having at least 2 heptad units. The artificial  
CC peptides spontaneously heterodimerise into a self-assembled protein-based  
CC structural material. The method is used to produce nanoscale structural  
CC materials designed for specific functions e.g. membranes, fibres, the  
CC absorbants, reactive materials, etc. A specific application is in the  
CC fabrication of selectively permeable membranes for protective gloves and  
CC coated fabrics for fuel handlers and other hazardous materials. Materials  
CC for medical implants, wound healants and other medical treatments would  
CC also be feasible using this technology.

XX Sequence 42 AA;

Query Match 32.9%; Score 124; DB 19; Length 42;

Best Local Similarity 72.5%; Pred. No. 3.3e-06;

Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 16 GDLKNVAQLKRVSLKDKAAELKQVSRLENIEDLKA 55

DB 2 GDLXNVAVQLXRVSLXDXAAELXQVSRLEXNIEDLXA 41

RESULT 12

AAB08364

ID AAB08364 standard; peptide; 43 AA.

XX AC AAB08364;

XX 20-DEC-2000 (first entry)

XX Amino acid sequence of a coiled-coil peptide.

XX Binding partner; protein modification; post-translational modification;

XX modulator; coiled-coil structure.

XX Unidentified.

XX Key Location/Qualifiers

XX Cleavage-site 24

XX /note= "thrombin cleavage site"

XX Misc-difference 39

XX /note= "fluorophore attachment site"

XX WO200050902-A2.

XX 31-AUG-2000.

XX 25-FEB-1999; 2000WO-GB00669.

XX 25-FEB-1999; 99GB-0004398.

XX (FLUO-) FLUORESCENCE LTD.

XX Colyer J, Craig RK, Maschio A, Mezna M;

XX WPI; 2000-572119/53.

PT High throughput assay for monitoring modification of polypeptides and  
PT modulation of the modifications -  
XX Disclosure; Page 61; 128pp; English.

XX The specification describes a method for analysing a sample. The method  
CC comprises immobilising a polypeptide to a physical support, contacting  
CC the immobilised polypeptide with a test sample which may contain an  
CC agent capable of modifying the immobilised polypeptide, contacting the  
CC immobilised polypeptide with a binding partner polypeptide, where  
CC association of both polypeptide is dependent on the modification  
CC state of the immobilised polypeptide, and measuring the association of  
CC the binding partner polypeptide to the immobilised polypeptide. The  
CC polypeptides, support and methods can be used to analyse a sample to  
CC determine if modification of a polypeptide is taking place and to  
CC identify modulators of the modification. This is useful for monitoring  
CC the post-translational modification of proteins. AAB08364-65 represent  
CC binding partners which have coiled-coil structures, and may be used in  
CC the method of the invention to assay for thrombin activity.

XX Sequence 43 AA;

Query Match 32.1%; Score 121; DB 21; Length 43;

Best Local Similarity 46.7%; Pred. No. 7e-06;

Matches 21; Conservative 14; Mismatches 8; Indels 2; Gaps 1;

QY 5 HHHHHGSMASGDLKNKVAQLKRVSLKDKAAELKQVSRLENE 49

DB 1 HHHHHGGGIA--QLEQEIAGLEQENRQLEQEIAGLEQEIARKEQ 43

RESULT 13

AAB08383

ID AAB08383 standard; peptide; 43 AA.

XX AC AAB08383;

XX 20-DEC-2000 (first entry)

XX Peptide used to assay for tobacco etch virus protease activity.

XX Binding partner; protein modification; post-translational modification;

XX modulator; coiled-coil structure.

XX Unidentified.

XX WO200050902-A2.

XX 31-AUG-2000.

XX 25-FEB-2000; 2000WO-GB00669.

XX 25-FEB-1999; 99GB-0004398.

XX (FLUO-) FLUORESCENCE LTD.

XX Colyer J, Craig RK, Maschio A, Mezna M;

XX WPI; 2000-572119/53.

XX High throughput assay for monitoring modification of polypeptides and  
XX modulation of the modifications -  
XX Example 9; Page 80; 128pp; English.

XX The specification describes a method for analysing a sample. The method  
CC comprises immobilising a polypeptide to a physical support, contacting  
CC the immobilised polypeptide with a test sample which may contain an  
CC agent capable of modifying the immobilised polypeptide, contacting the  
CC immobilised polypeptide with a binding partner polypeptide, where  
CC association of both polypeptide is dependent on the modification  
CC state of the immobilised polypeptide, and measuring the association of  
CC the binding partner polypeptide to the immobilised polypeptide. The





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:14:14 ; Search time 1.70376 Seconds  
(without alignments)  
1277.936 Million cell updates/sec

Title: US-09-490-291-11

Perfect score: 377

Sequence: 1 MRGSHHHHHGSMASGDLKN.....AKIGDLNNTSGIRPAKLN 74

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 2942292 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/PCTUS.COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	279.5	74.1	76	3	US-08-956-307B-18
2	243.5	64.6	76	3	US-08-956-307B-17
3	221	58.6	84	1	US-08-452-592B-7
4	213	56.5	84	1	US-08-452-592B-8
5	205	54.4	84	1	US-08-452-592B-9
6	199	52.8	42	3	US-08-956-307B-2
7	197	52.3	84	1	US-08-452-592B-10
8	181	48.0	84	1	US-08-452-592B-6
9	159	42.2	42	3	US-08-956-307B-1
10	124	32.9	42	1	US-08-452-592B-11
11	101.5	26.9	286	4	US-09-360-017-1
12	96	25.5	175	4	US-09-329-884-18
13	94	24.9	168	4	US-09-329-884-20
14	94	24.9	169	4	US-09-329-884-12
15	94	24.9	175	4	US-09-329-884-16
16	92	24.4	159	4	US-08-856-253-2
17	91.5	24.3	168	4	US-09-329-884-22
18	91.5	24.3	169	4	US-09-329-884-14
19	91	24.1	18	3	US-08-956-307B-4
20	90	23.9	345	4	US-08-856-253-7
21	87	23.1	14	3	US-08-956-307B-3
22	82.5	21.9	110	1	US-08-434-705B-15
23	82.5	21.9	110	2	US-09-086-201-15
24	82	21.8	18	1	US-08-578-649-14
25	81	21.5	42	5	PCT-US94-06655-9
26	81	21.5	48	1	US-08-457-245-21
27	81	21.5	84	4	US-09-299-495F-11

28	81	21.5	126	1	US-08-268-348A-12	Sequence 12, Appl
29	81	21.5	133	1	US-08-268-348A-8	Sequence 8, Appl
30	81	21.5	133	1	US-08-268-348A-10	Sequence 10, Appl
31	80.5	21.4	202	4	US-09-134-001C-5383	Sequence 5383, Ap
32	80.5	21.4	1587	4	US-09-000-094-46	Sequence 46, Appl
33	79.5	21.1	31	2	US-08-662-227-37	Sequence 37, Appl
34	79.5	21.1	31	4	US-09-017-947-37	Sequence 37, Appl
35	79	21.0	32	3	US-08-737-336-1	Sequence 1, Appl
36	79	21.0	39	2	US-08-679-865-39	Sequence 39, Appl
37	79	21.0	39	2	US-08-680-876-39	Sequence 39, Appl
38	79	21.0	39	4	US-09-263-975-39	Sequence 39, Appl
39	79	21.0	69	2	US-08-687-865A-21	Sequence 21, Appl
40	79	21.0	69	4	US-09-043-711-21	Sequence 21, Appl
41	79	21.0	88	2	US-08-690-011A-11	Sequence 11, Appl
42	79	21.0	178	4	US-09-183-841-2	Sequence 2, Appl
43	79	21.0	211	4	US-08-856-253-4	Sequence 4, Appl
44	79	21.0	254	4	US-09-004-731-89	Sequence 89, Appl
45	79	21.0	254	4	US-09-004-731-92	Sequence 92, Appl

ALIGNMENTS

RESULT 1  
US-08-956-307B-18  
; Sequence 18, Application US/08956307B  
; Patent No. 6090911  
; GENERAL INFORMATION:  
; APPLICANT: Petka, Wendy A.  
; APPLICANT: Tirrell, David A.  
; APPLICANT: Kevin P. McGrath  
; TITLE OF INVENTION: REVERSIBLE HYDROGELS  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: FastSeq for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; FILING DATE: 22-OCT-1997  
; APPLICATION NUMBER: US/08/956,307B  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fasse, J. Peter  
; REGISTRATION NUMBER: 32,983  
; REFERENCE/DOCKET NUMBER: 07880/033001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 76 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-956-307B-18

Query Match 74.1%; Score 279.5; DB 3; Length 76;

Best Local Similarity 89.1%; Pred. No. 2.3e-23;

Matches 57; Conservative 1; Mismatches 1; Indels 5; Gaps 1;

QY 1 MRGSHHHHHGSMASGDLKNVAOLKRVSLKDKAAELKQEVSRLENEEDLKA 55  
|||||

```
Db 1 MRGSHHHHGGDDDDKASGDLKNKVAQLKRVSLKDKAAELKQVSRSLKNEIEDLKA 60
QY 56 KIGD 59
    ||||
Db 61 KIGD 64

RESULT 2
US-08-956-307B-17
; Sequence 17, Application US/08956307B
; Patent No. 6090911
; GENERAL INFORMATION:
; APPLICANT: Pecka, Wendy A.
; APPLICANT: Tirrell, David A.
; APPLICANT: Kevin P. McGrath
; TITLE OF INVENTION: REVERSIBLE HYDROGELS
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,307B
; FILING DATE: 22-OCT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07880/033001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 76 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-956-307B-17

Query Match 64.6%; Score 243.5; DB 3; Length 76;
Best Local Similarity 75.0%; Pred. No. 1.7e-19;
Matches 48; Conservative 10; Mismatches 1; Indels 5; Gaps 1;

QY 1 MRGSHHHHGG-----MASGDLKNKVAQLKRVSLKDKAAELKQVSRSLKNEIEDLKA 55
    |||||
Db 1 MRGSHHHHGGDDDDKASGDLKNEVAQLEREVSRSLKNEIEDLKA 60

QY 56 KIGD 59
    ||||
Db 61 EIGD 64

RESULT 3
US-08-452-592B-7
; Sequence 7, Application US/08452592B
; Patent No. 5712366
; GENERAL INFORMATION:
; APPLICANT: McGrath, Kevin P.
; APPLICANT: Kaplan, David L.
; TITLE OF INVENTION: Fabrication of Nanoscale Materials Using
; FILING DATE: 25 May 1995
; CLASSIFICATION: 530
```

```
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: U.S. Army Soldier Systems Command
; STREET: AMSCC-CC (Patent Counsel)
; CITY: Natick
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01760-5035
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: DOS 6.22
; SOFTWARE: WordPerfect Version 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,592B
; FILING DATE: 25 May 1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/068,948
; FILING DATE: 25 May 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Donahue, Richard J.
; NAME: Lamming, John H.
; NAME: Ranucci, Vincent J.
; REGISTRATION NUMBER: 22,062 (Donahue)
; REGISTRATION NUMBER: 34,857 (Lamming)
; REGISTRATION NUMBER: 29,579 (Ranucci)
; REFERENCE/DOCKET NUMBER: NA-1096D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 508-233-4510
; TELEFAX: 508-233-5167
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-452-592B-7

Query Match 58.6%; Score 221; DB 1; Length 84;
Best Local Similarity 97.9%; Pred. No. 4.9e-17;
Matches 46; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 16 GDLKNKVAQLKRVSLKDKAAELKQVSRSLKNEIEDLKAIGDLNN 62
    |||||
Db 2 GDLKNKVAQLKRVSLKDKAAELKQVSRSLKNEIEDLKAIGDLKN 48

RESULT 4
US-08-452-592B-8
; Sequence 8, Application US/08452592B
; Patent No. 5712366
; GENERAL INFORMATION:
; APPLICANT: McGrath, Kevin P.
; APPLICANT: Kaplan, David L.
; TITLE OF INVENTION: Fabrication of Nanoscale Materials Using
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: U.S. Army Soldier Systems Command
; STREET: AMSCC-CC (Patent Counsel)
; CITY: Natick
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01760-5035
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: DOS 6.22
; SOFTWARE: WordPerfect Version 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,592B
; FILING DATE: 25 May 1995
; CLASSIFICATION: 530
```

```
/
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/068,948
/ FILING DATE: 25 May 1993
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Donahue, Richard J.
/ NAME: Lanning, John H.
/ NAME: Ranucci, Vincent J.
/ REGISTRATION NUMBER: 22,062 (Donahue)
/ REGISTRATION NUMBER: 34,857 (Lanning)
/ REGISTRATION NUMBER: 29,579 (Ranucci)
/ REFERENCE/DOCKET NUMBER: NA-1096D
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 508-233-4510
/ TELEFAX: 508-233-5167
/ INFORMATION FOR SEQ ID NO: 8:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 84 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ US-08-452-592B-8

Query Match 56.5%; Score 213; DB 1; Length 84;
Best Local Similarity 93.6%; Pred. No. 3.6e-16;
Matches 44; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 16 GDLKNKVAQLKRRVSLKDKAAELKQEVSRLENEIEDLKAKIGDLNN 62
Db 2 GDLKNKVAQLEREVSRSLKDKAAELKQEVSRLENEIEDLKAKIGDLKN 48

RESULT 5
US-08-452-592B-9
/ Sequence 9, Application US/08452592B
/ Patent No. 5712366
/ GENERAL INFORMATION:
/ APPLICANT: McGrath, Kevin P.
/ APPLICANT: Kaplan, David L.
/ TITLE OF INVENTION: Fabrication of Nanoscale Materials Using
/ TITLE OF INVENTION: Self-Assembling Proteins
/ NUMBER OF SEQUENCES: 11
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: U.S. Army Soldier Systems Command
/ STREET: AMSCC-CC (Patent Counsel)
/ CITY: Natick
/ STATE: Massachusetts
/ COUNTRY: United States of America
/ ZIP: 01760-5035
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
/ COMPUTER: IBM PC Compatible
/ OPERATING SYSTEM: DOS 6.22
/ SOFTWARE: WordPerfect Version 6.1
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/452,592B
/ FILING DATE: 25 May 1995
/ CLASSIFICATION: 530
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/068,948
/ FILING DATE: 25 May 1993
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Donahue, Richard J.
/ NAME: Lanning, John H.
/ NAME: Ranucci, Vincent J.
/ REGISTRATION NUMBER: 22,062 (Donahue)
/ REGISTRATION NUMBER: 34,857 (Lanning)
/ REGISTRATION NUMBER: 29,579 (Ranucci)
/ REFERENCE/DOCKET NUMBER: NA-1096D
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 508-233-4510
/ TELEFAX: 508-233-5167
/ INFORMATION FOR SEQ ID NO: 9:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 84 amino acids
```

```
/
/ TYPE: amino acid
/ TOPOLOGY: linear
/ US-08-452-592B-9

Query Match 54.4%; Score 205; DB 1; Length 84;
Best Local Similarity 89.4%; Pred. No. 2.6e-15;
Matches 42; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 16 GDLKNKVAQLKRRVSLKDKAAELKQEVSRLENEIEDLKAKIGDLNN 62
Db 2 GDLNEVAQLEREVSRSLKDKAAELKQEVSRLENEIEDLKAKIGDLN 48

RESULT 6
US-08-956-307B-2
/ Sequence 2, Application US/08956307B
/ Patent No. 6090911
/ GENERAL INFORMATION:
/ APPLICANT: Petka, Wendy A.
/ APPLICANT: Tirrell, David A.
/ APPLICANT: Kevin P. McGrath
/ TITLE OF INVENTION: REVERSIBLE HYDROGELS
/ NUMBER OF SEQUENCES: 24
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Fish & Richardson P.C.
/ STREET: 225 Franklin Street
/ CITY: Boston
/ STATE: MA
/ COUNTRY: USA
/ ZIP: 02110-2804
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: Windows 95
/ SOFTWARE: FastSeq for Windows Version 2.0b
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/956,307B
/ FILING DATE: 22-OCT-1997
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Fasse, J. Peter
/ REGISTRATION NUMBER: 32,983
/ REFERENCE/DOCKET NUMBER: 07880/033001
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 617/542-5070
/ TELEFAX: 617/542-8906
/ TELEX: 200154
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 42 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ US-08-956-307B-2

Query Match 52.8%; Score 199; DB 3; Length 42;
Best Local Similarity 100.0%; Pred. No. 5e-15;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 SGDLKNKVAQLKRRVSLKDKAAELKQEVSRLENEIEDLKAK 56
Db 1 SGDLKNKVAQLKRRVSLKDKAAELKQEVSRLENEIEDLKAK 42

RESULT 7
US-08-452-592B-10
/ Sequence 10, Application US/08452592B
/ Patent No. 5712366
/ GENERAL INFORMATION:
/ APPLICANT: McGrath, Kevin P.
/ APPLICANT: Kaplan, David L.
```



```

; APPLICANT: Heinrikson, Robert I.
; APPLICANT: Tomasselli, Alfredo G.
; APPLICANT: Koepfinger, Kenneth A.
; TITLE OF INVENTION: Method for Autoactivation of Procaspase 8
; FILE REFERENCE: Docket No. 6303329 6172 N CP
; CURRENT APPLICATION NUMBER: US/09/360,017
; CURRENT FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-360-017-1

Query Match      26.9%; Score 101.5; DB 4; Length 286;
Best Local Similarity 34.6%; Pred. No. 0.0014;
Matches 27; Conservative 8; Mismatches 20; Indels 23; Gaps 3;

QY 1 MRGSHHHHHHGSMAAGD-----LKNKVAQLKKRVRSLSL-----KDKAALKEVSR 45
   ||||| ||||| | :|| :| | | :| :| | | :| :| | | :| :| |
Db 1 MRGSHHHHHHGMTSDSPREDSQESOTLDKVYQMKSPPRGYCIIINHNFAKAREKVPK 60
   ||||| ||||| | :|| :| | | :| :| | | :| :| | | :| :| |

QY 45 LENEIEDLKAKIGDLNNT 53
   | :| :| | | | | |
Db 61 LHS-----IRDRNGT 70

RESULT 12
US-09-329-884-18
; Sequence 18, Application US/09329884
; Patent No. 6342233
; GENERAL INFORMATION:
; APPLICANT: Irvin, Randall T.
; APPLICANT: Hodges, Robert S.
; TITLE OF INVENTION: PSEUDOMONAS TREATMENT
; TITLE OF INVENTION: COMPOSITION AND METHOD
; FILE REFERENCE: 8900-0008.30
; CURRENT APPLICATION NUMBER: US/09/329,884
; CURRENT FILING DATE: 1999-06-11
; EARLIER APPLICATION NUMBER: US 60/089,155
; EARLIER FILING DATE: 1998-06-12
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-329-884-18

Query Match      25.5%; Score 96; DB 4; Length 175;
Best Local Similarity 28.6%; Pred. No. 0.0031;
Matches 22; Conservative 15; Mismatches 24; Indels 16; Gaps 1;

QY 5 HHHHHHGSMAAGDLKNKVAQLKKRVRSLSLKDKAELKEVSRLENE----- 49
   ||||| | :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 4 HHHHGGGGVSALEREVSALEREVSALEREVSALEREVSALEREVSALEREVSALEREVS 63
   || :| :| :| :| :| :| :| :| :| :| :| :| :| :|

QY 50 -IEDLKAKIGDLNNTSG 65
   || :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 64 LASGLTKVSDIFSQQG 80

RESULT 13
US-09-329-884-20
; Sequence 20, Application US/09329884
; Patent No. 6342233
; GENERAL INFORMATION:
; APPLICANT: Irvin, Randall T.
; APPLICANT: Hodges, Robert S.
; TITLE OF INVENTION: PSEUDOMONAS TREATMENT
; TITLE OF INVENTION: COMPOSITION AND METHOD
; FILE REFERENCE: 8900-0008.30
```



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:21:10 ; Search time 1.02236 Seconds  
(without alignments)  
1213.493 Million cell updates/sec

Title: US-09-490-291-11  
Perfect score: 377  
Sequence: 1 MRGSHHHHHGSMASGDLKN.....AKIGDLNNTSGIRPAKLN 74

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 106657 seqs, 16763532 residues

Total number of hits satisfying chosen parameters: 106657

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep:\*
- 7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB pep:\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep:\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB pep:\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB pep:\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB pep:\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB pep:\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB pep:\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	104	27.6	131	9	US-09-982-992A-2
2	101.5	26.9	286	10	US-09-862-915-1
3	96	25.5	175	10	US-09-865-159-18
4	95.5	25.3	128	10	US-09-742-373-9
5	94	24.9	168	10	US-09-865-159-20
6	94	24.9	169	10	US-09-865-159-12
7	94	24.9	175	10	US-09-865-159-16
8	93.5	24.8	323	9	US-09-987-107-58
9	92	24.4	159	10	US-09-813-820-2
10	92	24.4	336	9	US-09-987-107-44
11	92	24.4	337	9	US-09-987-107-46
12	91.5	24.3	168	10	US-09-865-159-22
13	91.5	24.3	169	10	US-09-865-159-14
14	91.5	24.3	324	12	US-10-007-805-551
15	90	23.9	273	9	US-09-987-107-50
16	90	23.9	345	10	US-09-813-820-7
17	89.5	23.7	323	9	US-09-987-107-56
18	88	23.3	324	9	US-09-987-107-62
19	88	23.3	324	9	US-09-987-107-64

20	87.5	23.2	323	9	US-09-987-107-60
21	87	23.1	265	10	US-09-732-091-44
22	83.5	22.1	316	9	US-09-987-107-48
23	83	22.0	504	10	US-09-732-091-42
24	82	21.8	92	10	US-09-833-747A-2
25	82	21.8	324	9	US-09-987-107-66
26	81	21.5	316	9	US-09-987-107-54
27	80	21.2	104	9	US-09-870-759-90
28	79.5	21.1	31	10	US-09-925-442-37
29	79	21.0	39	10	US-09-884-681-39
30	79	21.0	211	10	US-09-813-820-4
31	78	20.7	139	10	US-09-813-820-8
32	78	20.7	439	9	US-10-115-984-2
33	78	20.7	512	10	US-09-813-820-6
34	77	20.4	559	10	US-09-981-649A-28
35	76.5	20.3	14	9	US-09-784-199-9
36	76.5	20.3	193	9	US-09-991-496-119
37	76.5	20.3	193	10	US-09-874-923-119
38	76	20.2	326	9	US-09-991-211-1
39	75	19.9	30	10	US-09-754-826-3
40	74.5	19.8	261	9	US-09-987-107-52
41	74.5	19.8	344	9	US-09-987-107-68
42	74.5	19.8	716	10	US-09-845-157-2
43	74.5	19.8	879	9	US-10-108-605-217
44	74	19.6	17	9	US-09-051-013-6
45	74	19.6	20	9	US-09-051-013-3

## ALIGNMENTS

RESULT 1  
US-09-982-992A-2  
; Sequence 2, Application US/09982992A  
; Patent No. US20020164337A1  
; GENERAL INFORMATION:  
; APPLICANT: PATTI, Joseph M. et al.  
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES TO THE MAP PROTEIN AND METHOD OF USE IN  
; TITLE OF INVENTION: AND PREVENTING INFECTIONS  
; FILE REFERENCE: P06922US02/BAS  
; CURRENT APPLICATION NUMBER: US/09/982.992A  
; PRIOR FILING DATE: 2001-10-22  
; PRIOR APPLICATION NUMBER: 60/277,287  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/241,832  
; PRIOR FILING DATE: 2000-10-20  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 131  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-982-992A-2

Query Match	27.6%	Score 104;	DB 9;	Length 131;
Best Local Similarity	30.7%	Pred. No. 0.0029;		
Matches	31;	Conservative	9;	Mismatches 27; Indels 34; Gaps 3;
Qy	1	MRGSHHHHHGSG	-----	MASGDLKNKVAQLKRVRLSKD 34
Db	1	MRGSHHHHHGSGIPYITVTNGTSQNLSSLTFTFNKQOISYKDIENKVKSVLYFNRGISD 60		
Qy	35	-----AAELKQEVSRLENEIEDLKAKI--GDLNNTSGIR 67		
Db	61	IDRLRSQAKYTVHFNGTKRVRDLKAGIHTADLINTSDIK 101		
RESULT 2				
US-09-862-915-1				
; Sequence 1, Application US/09862915				
; Patent No. US20020045205A1				
; GENERAL INFORMATION:				
; APPLICANT: Heinrichson, Robert I.				

```
; APPLICANT: Tomasselli, Alfredo G.
; TITLE OF INVENTION: Method for Autoactivation of Procaspase 8
; FILE REFERENCE: Docket No. US20020045205A1 6172
; CURRENT APPLICATION NUMBER: US/09/862,915
; CURRENT FILING DATE: 2001-05-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-862-915-1

Query Match      26.9%; Score 101.5; DB 10; Length 286;
Best Local Similarity 34.6%; Pred. No. 0.011;
Matches 27; Conservative 8; Mismatches 20; Indels 23; Gaps 3;

QY 1 MRGSHHHHHGSMASGD-----LKNKVAQLKKRVSL-----KDKAAELKQEVSR 45
Db 1 MRGSHHHHHGSMTISDSPREODSESQTLDKVYQMKSPRGYCLIIINNNHFAKAREKVPK 60

QY 46 LENEIEDLKAIGDLNNT 63
Db 61 LHS-----INRNGT 70

RESULT 3
US-09-865-159-18
; Sequence 18, Application US/09865159
; Patent No. US20020098196A1
; GENERAL INFORMATION:
; APPLICANT: Irvin, Randall T.
; APPLICANT: Hodges, Robert S.
; TITLE OF INVENTION: PSEUDOMONAS TREATMENT
; FILE REFERENCE: 8900-0008.30
; CURRENT APPLICATION NUMBER: US/09/865,159
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/329,884
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-865-159-18

Query Match      25.5%; Score 96; DB 10; Length 175;
Best Local Similarity 28.6%; Pred. No. 0.021;
Matches 22; Conservative 15; Mismatches 24; Indels 16; Gaps 1;

QY 5 HHHHHGSMASGDLKNKVAQLKKRVSLKDKAAELKQEVSRLENE----- 49
Db 4 HHHHGGGGEVSALEKEVSALEKEVSALEKEVSALEKGGGGBFARAQLSERMT 63

QY 50 -IEDLKAIGDLNNTSG 65
Db 64 LASGLTKVSDIFSDQG 80

RESULT 4
US-09-742-373-9
; Sequence 9, Application US/09742373
; Patent No. US20020052471A1
; GENERAL INFORMATION:
; APPLICANT: Althaus, Harald
; APPLICANT: Hauser, Hans-Peter
; TITLE OF INVENTION: Human Procalcitonin and the Preparation and Use Thereof
; FILE REFERENCE: 05552.1445-00
; CURRENT APPLICATION NUMBER: US/09/742,373
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 19962434.8
```

```
; PRIOR FILING DATE: 1999-12-22
; PRIOR APPLICATION NUMBER: 10016278.9
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: 10027954.6
; PRIOR FILING DATE: 2000-06-08
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 128
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Protein, human
US-09-742-373-9

Query Match      25.3%; Score 95.5; DB 10; Length 128;
Best Local Similarity 30.8%; Pred. No. 0.017;
Matches 24; Conservative 18; Mismatches 17; Indels 19; Gaps 3;

QY 1 MRGSHHHHHGSG-----MASGDLKNKVAQLKKRVSLKDKAAELKQEVSR 44
Db 1 MRGSHHHHHGSAFFRSALLESSPADPATLSEARLRLAALVQDY--VOMKASELEQEQE 58

QY 45 RLENEIEDLKAIGDLN 61
Db 59 REGSSLDSPSRKRCGNLS 76

RESULT 5
US-09-865-159-20
; Sequence 20, Application US/09865159
; Patent No. US20020098196A1
; GENERAL INFORMATION:
; APPLICANT: Irvin, Randall T.
; APPLICANT: Hodges, Robert S.
; TITLE OF INVENTION: PSEUDOMONAS TREATMENT
; FILE REFERENCE: 8900-0008.30
; CURRENT APPLICATION NUMBER: US/09/865,159
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/329,884
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 20
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-865-159-20

Query Match      24.9%; Score 94; DB 10; Length 168;
Best Local Similarity 40.0%; Pred. No. 0.03;
Matches 20; Conservative 8; Mismatches 16; Indels 6; Gaps 1;

QY 6 HHHHHGSMASGDLKNKVAQLKKRVSLKDKAAELKQEVSRLENEIEDLKA 55
Db 4 HHHHHGSG-----GGEIEALKAEIEALKAEIEALKAEIEALKAEIEALK 47

RESULT 6
US-09-865-159-12
; Sequence 12, Application US/09865159
; Patent No. US20020098196A1
; GENERAL INFORMATION:
; APPLICANT: Irvin, Randall T.
; APPLICANT: Hodges, Robert S.
; TITLE OF INVENTION: PSEUDOMONAS TREATMENT
; FILE REFERENCE: 8900-0008.30
; CURRENT APPLICATION NUMBER: US/09/865,159
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/329,884
```



```
QY 1 MRGSHHHHHHGGSMASGDLKNKV 22
      |||||
Db 1 MRGSHHHHHHGGSITSGNKSTNV 22
```



```

; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Durham, Margarita
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.470C10
; CURRENT APPLICATION NUMBER: US/10/007,805
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 593
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 551
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-007-805-551

Query Match      24.3%; Score 91.5; DB 12; Length 324;
Best Local Similarity 32.4%; Pred. No. 0.1;
Matches 24; Conservative 11; Mismatches 24; Indels 15; Gaps 2;

QY 5 HHHHHHGSMA-----SGDLKNKVAQLKRRKRSKDKAAELKQEVSRLEN---E 49
   |||||:| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3 HHHHHHGTALQCEVSHTHENENYLLHENCMLKKEIAMLKLEIATLKHQYQEKENKYFED 62

QY 50 IEDLKAKIGDLNNT 63
   | : | | | : | |
Db 63 IKILKEKNAELQMT 76

RESULT 15
US-09-987-107-50
; Sequence 50, Application US/09987107
; Patent No. US20020156007A1
; GENERAL INFORMATION:
; APPLICANT: GRAVERSEN, Jonas
; APPLICANT: MOESTRUP, Soren
; TITLE OF INVENTION: APOLIPOPROTEINS ANALOGUES
; FILE REFERENCE: GRAVERSEN1A
; CURRENT APPLICATION NUMBER: US/09/987,107
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/264,022
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: DK PA2001 00057
; PRIOR FILING DATE: 2001-01-15
; PRIOR APPLICATION NUMBER: DK PA2000 01682
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: pT7H6 Trip-A-Apo A-1-del 43 - AmpR plasmid
US-09-987-107-50

Query Match      23.9%; Score 90; DB 9; Length 273;
Best Local Similarity 29.2%; Pred. No. 0.12;
Matches 26; Conservative 11; Mismatches 34; Indels 18; Gaps 3;

QY 3 GSHHHHHHGSMA-----GDLKNKVAQLKRRK-----RSLKAAELKQEVSRLEN 48
   |||||:| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2 GSHHHHHHGSIQGRSPGTEPTQPKKIVNAKDVNTKMFEEKLSRLDTLAEVALLKE 61
```

```

QY 49 ----EIEDLKAKIGDLNNTSGIRRPAAKL 73
   || : | : | : | | : | |
Db 62 QQALQTVSLKGSCLKLLDNWDSVTSTFSKL 90

Search completed: December 18, 2002, 16:39:47
Job time : 17.0223 secs
```

***This Page Blank (uspio)***

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:13:49 ; Search time 2.18649 Seconds  
(without alignments)  
3253.588 Million cell updates/sec

Title: US-09-490-291-11  
Perfect score: 377  
Sequence: 1 MRCSSHHHHHSGMASGDLKN.....AKIGDLNNTSGIRRAKLN 74

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR.73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	94.5	25.1	1186	2 G69708	chromosome segrega
2	89	23.6	315	2 JC7572	somite Waf1 protei
3	88	23.3	259	2 A70359	hydrogenase expres
4	86	22.8	155	2 I50167	B-G antigen - chic
5	85.5	22.7	323	2 I49529	transcription fact
6	85	22.5	321	2 D39371	Ig V-region-like B
7	85	22.5	880	2 F75103	conserved hypothet
8	85	22.5	979	2 JQ0894	155 protein - Myc
9	84.5	22.4	281	2 F75216	hypothetical prote
10	84	22.3	1199	2 T29145	hypothetical prote
11	82.5	21.9	397	2 F90182	hypothetical prote
12	82	21.8	173	2 JC5610	troponin I - sea s
13	82	21.8	311	2 A56235	transcription acti
14	82	21.8	766	2 S37894	hypothetical prote
15	82	21.8	1188	2 G83960	chromosome segrega
16	81.5	21.6	212	2 H69402	hypothetical prote
17	81	21.5	166	2 S73342	hypothetical prote
18	81	21.5	270	2 G71061	hypothetical prote
19	81	21.5	396	2 T41405	hypothetical repea
20	81	21.5	409	2 F75204	probable 26S prote
21	80.5	21.4	269	2 T18335	icmG protein - Leg
22	80.5	21.4	647	2 A84265	hypothetical prote
23	80	21.2	102	2 F64341	hypothetical prote
24	80	21.2	103	2 B60608	myosin heavy chain
25	80	21.2	527	2 S33068	myosin heavy chain
26	80	21.2	1225	2 A49464	chromosome segrega
27	80	21.2	1940	2 A59287	myosin heavy chain
28	79.5	21.1	565	1 HMIVE3	hemagglutinin prec
29	79.5	21.1	565	1 HMIVE4	hemagglutinin prec

30	79	21.0	318	2 T49167	hypothetical prote
31	79	21.0	369	1 TVFVAF	transforming prote
32	79	21.0	407	2 F72343	hypothetical prote
33	79	21.0	554	2 G72361	hypothetical prote
34	78.5	20.8	564	1 HMIVE2	hemagglutinin prec
35	78.5	20.8	565	1 HMIVE2	hemagglutinin prec
36	78.5	20.8	1956	2 T16416	hypothetical prote
37	78	20.7	456	2 E86903	hypothetical prote
38	78	20.7	1169	2 A64505	P115 homolog - Met
39	78	20.7	1313	2 A48467	myosin heavy chain
40	78	20.7	1938	1 A40997	myosin heavy chain
41	77.5	20.6	348	2 T04618	heat shock protein
42	77.5	20.6	778	2 T30430	hypothetical prote
43	77.5	20.6	1556	2 F96587	hypothetical prote
44	77	20.4	292	2 I51171	transcription fact
45	77	20.4	338	2 I38567	retrovirus-related

## ALIGNMENTS

### RESULT 1

G69708

chromosome segregation SMC protein - Bacillus subtilis  
N:Alternate names: minichromosome stabilizing protein SMC  
C:Species: Bacillus subtilis

C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
C:Accession: G69708; JC4819; PC4029  
R:Kunst, F.; Ogawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber  
C.; Bron, S.; Brouillette, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,  
Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal  
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.  
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau  
Y. M.; Ogawa, K.; Ogawara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete  
Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili  
A:Reference number: A69580; MUID:9804033; PMID:9384377

A:Accession: G69708  
A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1186 <KUN>

A:Cross-references: GB:299112; GB:AL009126; NID:g2633902; PIDN:CAB13467.1; PID:g26339

R:Oguro, A.; Kakeshita, H.; Takamatsu, H.; Nakamura, K.; Yamane, K.

A:Title: The effect of Srb, a homologue of the mammalian SRP receptor alpha-subunit,

Gene 172, 17-24, 1996

A:Reference number: JC4819; MUID:96257247; PMID:8654983

A:Accession: JC4819

A:Molecule type: DNA

A:Residues: 1-49, 'G', '51-161, 'G', '163-174, 'E', '176-177, 'G', '179-191, 'G', '193-227, 'P', '229-2  
74, 'P', '476-493, 'D', '495-514, 'V', '516-541, 'D', '547-585, 'SKLRGNSGPAFIISF', '601  
, 'F', '741-1186 <OGU>

A:Cross-references: DDBJ:D64116; NID:g1389548; PIDN:BAAL0577.1; PID:g1237015  
R:Oguro, A.; Kakeshita, H.; Honda, K.; Takamatsu, H.; Nakamura, K.; Yamane, K.

DNA Res. 2, 93-100, 1995

A:Title: srb: a Bacillus subtilis gene encoding a homologue of the alpha-subunit of t

A:Reference number: JC4093; MUID:96093930; PMID:7584053

A:Accession: PC4029

A:Molecule type: DNA

A:Residues: 1171-1186 <OG2>

A:Cross-references: DDBJ:D49781; NID:g1237017; PIDN:BAA08615.1; PID:g1237018

C:Genetics:

C:Superfamily: chromosome segregation protein SMC1

Query Match 25.1% Score 94.5; DB 2; Length 1186;



C:Species: Gallus gallus (chicken)  
C:Date: 20-Mar-1992 #sequence\_revision 10-Apr-1992 #text\_change 21-Jul-2000  
C:Accession: D39371  
R:Miller, M.M.; Goto, R.; Young, S.; Chirivella, J.; Hawke, D.; Miyada, C.G.  
Proc. Natl. Acad. Sci. U.S.A. 88, 4377-4381, 1991  
A:Title: Immunoglobulin variable-region-like domains of diverse sequence within the major histocompatibility complex class II genes of Gallus gallus  
A:Reference number: A39371; MUID:91239571; PMID:1903541  
A:Accession: D39371  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-321 <MIL>  
A:Cross-references: GB:M61862; NID:g211259; PIDN:AAA48627.1; PID:g211261

Query Match 22.5%; Score 85; DB 2; Length 321;  
Best Local Similarity 26.8%; Pred. No. 4.8;  
Matches 22; Conservative 18; Mismatches 24; Indels 18; Gaps 2;

Qy 7 HHGHSMAAGDLK-----NKVAQLKRVKSLDKAAELKQEVSRLENEIEDL 52  
Db 185 HLAEDLSTADLKLLAAKLVEQREAVERDQLKQYKELGSRATNLTKLLENEIEE 244

Qy 53 LKAKIGDLNNTSGIRRAAKLN 74  
Db 245 VEKHLKKI----GIRAPNKLH 262

RESULT 7  
F75103  
conserved hypothetical protein PAB0812 - Pyrococcus abyssi (strain Orsay)  
C:Species: Pyrococcus abyssi  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 02-Feb-2001  
C:Accession: F75103  
R:anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999  
A:Description: Pyrococcus abyssi genome sequence: Insights into archaeal chromosome structure  
A:Reference number: A75001  
A:Accession: F75103  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-880 <KAW>  
A:Cross-references: GB:AJ248286; GB:AL096836; NID:g5458366; PIDN:CAB50131.1; PID:g545864  
A:Experimental source: strain Orsay  
C:Genetics:  
A:Gene: PAB0812  
C:Superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF1032

Query Match 22.5%; Score 85; DB 2; Length 880;  
Best Local Similarity 37.2%; Pred. No. 13;  
Matches 16; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

Qy 18 LKNKVAQLKRVKSLDKAAELKQEVSRLENEIEDLAKIGDL 60  
Db 240 LKGISELKIQVEKLKGRKGLKEIKVQIERSIEKKAKISEL 282

RESULT 8  
JQ0894  
p115 protein - Mycoplasma hyorhinis  
C:Species: Mycoplasma hyorhinis  
C:Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 02-Feb-2001  
C:Accession: JQ0894  
R:Notarnicola, S.M.; McIntosh, M.A.; Wise, K.S.  
Gene 97, 77-85, 1991  
A:Reference number: JQ0894; MUID:91138990; PMID:1825306  
A:Title: A Mycoplasma hyorhinis protein with sequence similarities to nucleotide-binding domain of the ATP synthase  
A:Accession: JQ0894  
A:Molecule type: DNA  
A:Residues: 1-979 <NOT>  
A:Cross-references: GB:M34956  
A:Note: the authors translated the codon AAA for residue 956 as Leu  
C:Comment: This protein is located in the cytoplasm.  
C:Genetics:  
A:Genetic code: SGC3

C:Superfamily: chromosome segregation protein SMC1  
C:Keywords: nucleotide binding; P-loop  
F:32-39/Region: nucleotide-binding motif A (P-loop)

Query Match 22.5%; Score 85; DB 2; Length 979;  
Best Local Similarity 30.9%; Pred. No. 14;  
Matches 21; Conservative 15; Mismatches 22; Indels 10; Gaps 2;

Qy 17 DLKNKVAQLKRVKSL-----DKAAELKQEVSRLENEIEDLAKIGDLNNTSGI--- 66  
Db 323 DQKTKIEIKQVESLKQINASKQREIQLDQQLTRLNKAKANSLKQENDINKETGVILLE 382

Qy 67 RRPAAKLN 74  
Db 383 KKSAAAN 390

RESULT 9  
F75216  
hypothetical protein PAB2181 - Pyrococcus abyssi (strain Orsay)  
C:Species: Pyrococcus abyssi  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C:Accession: F75216  
R:anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999  
A:Description: Pyrococcus abyssi genome sequence: Insights into archaeal chromosome structure  
A:Reference number: A75001  
A:Accession: F75216  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-281 <KAW>  
A:Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB49181.1; PID:g5457454  
A:Experimental source: strain Orsay  
C:Genetics:  
A:Gene: PAB2181

Query Match 22.4%; Score 84.5; DB 2; Length 281;  
Best Local Similarity 40.8%; Pred. No. 4.6;  
Matches 20; Conservative 10; Mismatches 12; Indels 7; Gaps 1;

Qy 17 DLKNKVAQLKRVKSLDKAAELKQEVSRLENEIEDLAKIG 58  
Db 222 ELERKVSLESLNFEYFKVSLKKELEKNKVELEEVNKLKEGIG 270

RESULT 10  
T29145  
hypothetical protein F56A3.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T29145  
R:Geisler, C.; Bradshaw, H.; Wamsley, P.  
submitted to the EMBL Data Library, November 1996  
A:Description: The sequence of C. elegans cosmid F56A3.  
A:Reference number: T20578  
A:Accession: T29145  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1199 <GEI>  
A:Cross-references: EMBL:U80446; PIDN:AAB37802.1; GSPDB:GN00019; CESP:F56A3.4  
A:Experimental source: strain Bristol N2; clone F56A3  
C:Genetics:  
A:Gene: CESP:F56A3.4  
A:Map position: 1  
A:Introns: 18/1; 80/3; 446/3; 475/1; 607/3; 679/1; 974/2; 1058/1; 1168/3

Query Match 22.3%; Score 84; DB 2; Length 1199;  
Best Local Similarity 33.3%; Pred. No. 21;  
Matches 17; Conservative 13; Mismatches 21; Indels 0; Gaps 0;

Qy 12 SMASGDLNKNVAQLKRVKSLDKAAELKQEVSRLENEIEDLAKIGDLNN 62  
Db 198 TLTSGYEERKINDLEAKLLSEIDKVAELEDHIQURQELDDQSAARLASEN 248

```

RESULT 11
F90182
hypothetical protein SS00390 [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C:Accession: F90182
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, R.
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: F90182
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-397 <KUR>
A:CROSS-references: GB:AE006641; NID:g13813537; PIDN:AAK40717.1; GSPDB:GN00155
C:Genetics:
A:Gene: SS00390

Query Match 21.9%; Score 82.5; DB 2; Length 397;
Best Local Similarity 34.5%; Pred. No. 9.4;
Matches 19; Conservative 13; Mismatches 16; Indels 7; Gaps 1;

QY 18 LKNKVAQLKRRVSLDKAAELKQ-----EVSRLNEIEDLKAKIGDLNNTSG 65
Db 262 LNNEVSTLRSEISLSNSTIASLNKSLANANTQISNLQSEITTLNSEIGKLNSTVG 316

RESULT 12
JC5610
troponin I - sea squirt (Halocynthia roretzi)
C:Species: Halocynthia roretzi
C>Date: 23-Sep-1997 #sequence_revision 17-Oct-1997 #text_change 20-Jun-2000.
C:Accession: JC5610
R:Yuasa, H.J.; Sato, S.; Yamamoto, H.; Takagi, T.
J. Biochem. 122, 374-380, 1997
A:Title: Primary structure of troponin I isoforms from the ascidian Halocynthia roretzi.
A:Reference number: JC5610; MUID:98021076; PMID:9378716
A:Accession: JC5610
A:Molecule type: mRNA
A:Residues: 1-173 <YUA>
A:CROSS-references: DDBJ:AB001685; NID:g1888344; PIDN:BAA19425.1; PID:g1888345
A:Experimental source: adult Halocynthia roretzi
C:Comment: This protein binds to actin, and inhibits the interaction between actin and
C:Superfamily: troponin I

Query Match 21.8%; Score 82; DB 2; Length 173;
Best Local Similarity 30.5%; Pred. No. 4.5;
Matches 18; Conservative 16; Mismatches 23; Indels 2; Gaps 1;

QY 11 GSNASGDLKNKVAQLKRRVSLDKAAELKQEVSRLENEIEDLKAKIGDLNNTSGTRRP 69
Db 47 GGMSEQLDKLCRELHAKIEVDQRYDIEVKVKNQDEIEDLNQRIFDLRGK--FKRP 103

RESULT 13
A56235
transcription activator MafB - chicken
C:Species: Gallus gallus (chicken)
C>Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 20-Jun-2000
C:Accession: A56235
R:Kataoka, K.; Fujiwara, K.T.; Noda, M.; Nishizawa, M.
Mol. Cell. Biol. 14, 7581-7591, 1994
A:Title: MafB, a new Maf family transcription activator that can associate with Maf and
A:Reference number: A56235; MUID:95021288; PMID:7935473
A:Accession: A56235
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-311 <KAT>
A:CROSS-references: GB:D28600; NID:g516723; PIDN:BAA05938.1; PID:g516724
```

```

C:Genetics:
A:Introns: #status absent
C:Superfamily: maf transforming protein; maf homology
C:Keywords: DNA binding; homodimer; leucine zipper
F:200-289/Domain: maf homology <MAF>

Query Match 21.8%; Score 82; DB 2; Length 311;
Best Local Similarity 21.7%; Pred. No. 8.1;
Matches 30; Conservative 12; Mismatches 20; Indels 76; Gaps 3;

QY 5 HHHHHHCSMAS-----C 16
Db 159 HHHHHHHQASPTFTSSSSSOQLQTSHQHPFSSSVEDRFDQLVSMGVREINRHLRG 218
QY 17 DLKNKVAQLKRRVSLDKD-----AAELKQEVSRLENE 49
Db 219 FTKDEVIRLXQKRRITLKNRGYAQSCRYKRVQKHHLENEKTLTIQQVEQLKQEVTRLARE 278
QY 50 IEDLKAKIGDLNNTSGIR 67
Db 279 RDAYKLKCKEKL-ASNGFR 295

RESULT 14
S37894
hypothetical protein YKL072w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YKL352
C:Species: Saccharomyces cerevisiae
C>Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 29-Oct-1999
C:Accession: S37894; S37897; S39170; S44515
R:Rasmussen, S.; von Wettstein, D.
submitted to the Protein Sequence Database, March 1994
A:Reference number: S37872
A:Accession: S37894
A:Molecule type: DNA
A:Residues: 1-766 <RAS>
A:CROSS-references: EMBL:Z28072; NID:g486100; PIDN:CAA81909.1; PID:g486101; MIPS:YKLO
A:Experimental source: strain S288C
R:Pohl, T.M.; Pohl, F.M.
submitted to the Protein Sequence Database, March 1994
A:Reference number: S37897
A:Accession: S37897
A:Molecule type: DNA
A:Residues: 1-557 <POH>
A:CROSS-references: EMBL:Z28072; MIPS:YKL072w
A:Experimental source: strain S288C
R:Rasmussen, S.W.
submitted to the EMBL Data Library, November 1993
A:Reference number: S39168
A:Accession: S39170
A:Molecule type: DNA
A:Residues: 1-766 <RA2>
A:CROSS-references: EMBL:X75780; NID:g433625; PIDN:CAA53402.1; PID:g433628
R:Rasmussen, S.W.
Yeast 10, 69-74, 1994
A:Title: Sequence of a 20.7 kb region of yeast chromosome XI includes the NUP100 gene
rp in addition to seven ORFs with weak or no significant similarity to known proteins
A:Reference number: S44513
A:Accession: S44515
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-766 <RA2>
A:CROSS-references: EMBL:X75780; NID:g433625; PIDN:CAA53402.1; PID:g433628
C:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1993
C:Genetics:
A:Gene: SGD:STB6
A:CROSS-references: SGD:S0001555; MIPS:YKL072w
A:Map position: 11L

Query Match 21.8%; Score 82; DB 2; Length 766;
Best Local Similarity 26.7%; Pred. No. 20;
Matches 20; Conservative 20; Mismatches 25; Indels 10; Gaps 2;
```



Search completed: December 18, 2002, 16:22:25  
Job time : 4.18649 secs

***This Page Blank (uspto)***

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 16:06:04 : Search time 1.13584 Seconds  
(without alignments)  
2702.183 Million cell updates/sec

Title: US-09-490-291-11  
Perfect score: 377  
Sequence: .1 MRGSHHHHGGSMASGLKN.....AKIGLNNNTGIRRAAKLN 74

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	94.5	25.1	1186	1 SMC_BACSU	P51834 bacillus su
2	86.5	22.9	323	1 MAFB_RAT	P45842 rattus norv
3	85	22.8	548	1 CEAK_ECOLI	O47502 escherichia
4	85.5	22.7	323	1 MAFB_HUMAN	Q9Y5Q3 homo sapien
5	85.5	22.7	323	1 MAFB_MOUSE	P54841 mus musculus
6	85	22.5	880	1 RA50_PYRAB	Q9UZC8 pyrococcus
7	85	22.5	979	1 P115_MYCHR	P41508 mycoplasma
8	82	21.8	766	1 STB6_YEAST	P36085 saccharomyc
9	82	21.8	882	1 RA50_PYRPU	P58301 pyrococcus
10	81	21.5	166	1 YB38_MYCPN	P75250 mycoplasma
11	81	21.5	409	1 PSMR_AERPE	Q9YAC7 aeropyrum p
12	80	21.2	102	1 Y334_METJA	O57780 methanococc
13	80	21.2	1225	1 SMC1_YEAST	P32908 saccharomyc
14	79.5	21.1	565	1 HEMA_IAHAL	P16994 influenza a
15	79.5	21.1	565	1 HEMA_IAHTO	P17000 influenza a
16	79	21.0	369	1 TMAF_AVIS4	P23091 avian muscu
17	78.5	20.8	564	1 HEMA_IAGRE	P19698 influenza a
18	78.5	20.8	565	1 HEMA_IAHMI	P15658 influenza a
19	78	20.7	442	1 VATC_DROME	Q9V7N5 drosophila
20	78	20.7	1169	1 SMC_METJA	O59037 methanococ
21	78	20.7	1938	1 MYS_AEQIR	P24733 aequipesten
22	78	20.7	4349	1 DYHC_FUSSO	P78716 fusarium so
23	77	20.4	461	1 US45_LACLC	P22865 lactococcus
24	76.5	20.3	741	1 BSG2_DROME	P11929 drosophila
25	75.5	20.0	550	1 HEMA_IADHK	P43257 influenza a
26	75.5	20.0	1509	1 MYSN_ACACA	P05659 acanthamoeb
27	75	19.9	715	1 CLPB_MYCPN	P75247 mycoplasma
28	75	19.9	1727	1 ALMI_SCHPO	Q9UTK5 schizosacch
29	74.5	19.8	830	1 PAML_YEAST	P37304 saccharomyc
30	74.5	19.8	879	1 MYSP_DROME	P35415 drosophila
31	74.5	19.8	879	1 RA50_PYRHO	O58687 pyrococcus
32	74	19.6	539	1 MYS3_PYDAT	P39922 hydra atten
33	74	19.6	1957	1 YD86_SCHPO	Q10411 schizosacch

## ALIGNMENTS

RESULT 1

SMC\_BACSU

ID	SMC_BACSU	STANDARD;	PRT;	1186 AA.
AC	P51834; O31735;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Chromosome partition protein smc.			
GN	SMC.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
OX	NCBI_TaxID=1423;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RL	Oguro A., Kakeshita H., Takamatsu H., Nakamura K., Yamane K.;			
RL	Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=98044033; PubMed=9384377;			
RA	Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,			
RA	Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,			
RA	Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,			
RA	Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,			
RA	Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,			
RA	Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,			
RA	Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,			
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,			
RA	Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,			
RA	Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,			
RA	Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,			
RA	Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,			
RA	Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,			
RA	Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,			
RA	Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,			
RA	Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,			
RA	Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,			
RA	Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,			
RA	Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,			
RA	Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,			
RA	Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,			
RA	Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,			
RA	Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,			
RA	Takuchi M., Tanakashi A., Tanaka T., Terpstra P., Tognoni A.,			
RA	Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,			
RA	Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,			
RA	Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,			
RA	Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;			
RT	"The complete genome sequence of the Gram-positive bacterium Bacillus			
RT	subtilis.";			
RT	Nature 390:249-256(1997).			
RL	[3]			
RN	SEQUENCE OF 1171-1186 FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=96093930; PubMed=7584053;			

RA Oguro A., Kakeshita H., Honda K., Takamatsu H., Nakamura K.,  
RA Yamane K.;  
RT "srb: a Bacillus subtilis gene encoding a homologue of the alpha-  
RT subunit of the mammalian signal recognition particle receptor.";  
RL DNA Res. 2:95-100(1995).  
RN [4]  
RP CHARACTERIZATION.  
RX MEDLINE=98367134; PubMed=9701812;  
RA Moriya S., Tsujikawa E., Hassan A.K., Asai K., Kodama T.,  
RA Ogasawara N.;  
RT "A Bacillus subtilis gene-encoding protein homologous to eukaryotic  
RT SMC motor protein is necessary for chromosome partitioning";  
RL Mol. Microbiol. 29:179-187(1998).  
RN [5]  
RP CHARACTERIZATION.  
RX MEDLINE=98241483; PubMed=9573042;  
RA Britton R.A., Lin D.C., Grossman A.D.;  
RT "Characterization of a prokaryotic SMC protein involved in chromosome  
RT partitioning";  
RL Genes Dev. 12:1254-1259(1998).  
CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN CHROMOSOME STRUCTURE AND  
CC PARTITIONING. ESSENTIAL FOR CHROMOSOME PARTITION.  
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D841116; BAA10977.1; -;  
DR EMBL; Z991112; CAB13467.1; -;  
DR EMBL; D49781; BAA08615.1; -;  
DR Subtilist; BG11538; smc  
DR InterPro: IPR003439; ABC\_transport.  
DR InterPro: IPR003405; SMC\_C.  
DR InterPro: IPR003395; SMC\_N.  
DR Pfam; PF02463; SMC\_N; 1.  
DR Pfam; PF02483; SMC\_C; 1.  
KW ATP-binding; Coiled coil; Complete proteome.  
FT NP\_BIND 31 38  
FT DOMAIN 244 481  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 664 943  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 989 1031  
FT COILED COIL (POTENTIAL).  
FT E -> G (IN REF. 1).  
FT E -> G (IN REF. 1).  
FT K -> E (IN REF. 1).  
FT E -> G (IN REF. 1).  
FT E -> G (IN REF. 1).  
FT A -> P (IN REF. 1).  
FT K -> E (IN REF. 1).  
FT A -> P (IN REF. 1).  
FT A -> P (IN REF. 1).  
FT D -> G (IN REF. 1).  
FT N -> D (IN REF. 1).  
FT D -> D (IN REF. 1).  
FT KEELSKQ -> TRRAFEA (IN REF. 1).  
FT Q -> H (IN REF. 1).  
FT E -> K (IN REF. 1).  
FT I -> F (IN REF. 1).  
FT A -> P (IN REF. 1).  
FT E -> D (IN REF. 1).  
FT E -> D (IN REF. 1).  
FT L -> V (IN REF. 1).  
FT A -> P (IN REF. 1).  
FT QSRDAETARHSFL -> SKPLRNSGPAFTISF (IN  
FT REF. 1).  
FT TVLTEDLK -> NRSYRGLKG (IN REF. 1).  
FT A -> S (IN REF. 1).  
FT 664 664  
FT 676 676  
FT 680 680  
FT 694 694  
FT A -> S (IN REF. 1).

FT CONFLICT 701 701 K -> Q (IN REF. 1).  
FT CONFLICT 726 726 L -> V (IN REF. 1).  
FT CONFLICT 738 740 LQV -> POF (IN REF. 1).  
SQ SEQUENCE 1186 AA; 135510 MW; 0163227A2BCA2B CRC64;  
Query Match 25.1%; Score 94.5; DB 1; Length 1186;  
Best Local Similarity 36.8%; Pred. No. 0.79;  
Matches 25; Conservative 11; Mismatches 21; Indels 11; Gaps 2;  
QY 11 GSNASGDLKKNVQAQLKRVKSLKD-----KAAELAQEVSRLNEETEDLKAKIGDL 60  
Db 658 GSGTGAIVKKNNSLLGRSRELEDTYKRLAEMEEKTALLEQEVTKLHKSIQDMKKLADL 717  
QY 61 NNT-SGIR 67  
Db 718 RETGEGLR 725  
RESULT 2  
MAFB\_RAT  
ID MAFB\_RAT STANDARD; PRT; 323 AA.  
AC P54842;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Transcription factor MafB (V-maf musculoaponeurotic fibrosarcoma  
DE oncogene homolog B) (Transcription factor MAF1).  
GN MAFB OR MAF1  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Wistar; TISSUE=Liver;  
RX MEDLINE=97190228; PubMed=9038383;  
RA Sakai M., Imaki J., Yoshida K., Ogata A., Matsushima-Hibaya Y.,  
RA Kuboki Y., Nishizawa M., Nishi S.;  
RT "Rat maf related genes: specific expression in chondrocytes, lens and  
RT spinal cord.";  
RL Oncogene 14:745-750(1997).  
CC -!- FUNCTION: Plays a pivotal role in regulating lineage-specific  
CC hematopoiesis by repressing Ets1-mediated transcription of  
CC erythroid-specific genes in myeloid cells (By similarity).  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY. MAF SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U56241; AAB50062.1; -;  
DR InterPro: IPR004826; TF\_Maf.  
DR InterPro: IPR004827; TF\_BZIP.  
DR Pfam; PF03131; bzip\_Maf; 1.  
DR SMART; SM00338; BRLZ; 1.  
DR Transcription regulation; Repressor; DNA-binding; Nuclear protein.  
FT DNA\_BIND 238 264  
FT BASIC MOTIF.  
FT DOMAIN 266 287  
FT LEUCINE-ZIPPER.  
FT DOMAIN 131 143  
FT POLY-HIS.  
FT DOMAIN 158 167  
FT POLY-HIS.  
FT DOMAIN 194 201  
FT POLY-ALA.  
SQ SEQUENCE 323 AA; 35792 MW; 6E386340D1F840A5 CRC64;  
Query Match 22.9%; Score 86.5; DB 1; Length 323;  
Best Local Similarity 23.1%; Pred. No. 1;  
Matches 34; Conservative 9; Mismatches 19; Indels 85; Gaps 4;  
QY 5 HHHHHH-----GSMAS----- 15





```

DR SMART: SW00382; AA; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 30 37 ATP (BY SIMILARITY).
FT DOMAIN 144 745 COILED COIL (POTENTIAL).
SQ SEQUENCE 880 AA; 103970 MW; FDB177EC7E026479 CRC64;

Query Match 22.5%; Score 85; DB 1; Length 880;
Best Local Similarity 37.2%; Pred. No. 3.7;
Matches 16; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY 18 LKNKVAOLKKVRSLDKAAELKQEVSRLENEIEDLKAKIGDL 60
    | : : : : | : : : : | : : : : | : : : : | : : : : |
DB 240 IKGKISELKQVEKLGKRGKLEKIVQIERISIEEKKAKISEL 282

RESULT 7
P115_MYCHR STANDARD; PRT; 979 AA.
ID P115_MYCHR
AC P41508;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE P115 protein.
OS Mycoplasma hyorhinitis.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
CX NCBI_TaxID=2109;
[1]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
MEDLINE=91138950; PubMed=1825306;
Notarnicola S.M., McIntosh M.A., Wise K.S.;
"A Mycoplasma hyorhinitis protein with sequence similarities to
nucleotide-binding enzymes.";
Gene 97:77-85(1991).
RL CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
CC -1- FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY. SIMILAR TO OTHER MYCOPLASMA
CC P115.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@lsb-sib.ch).
-----
ENBL: M34956; AAA25423.1; --
PIR: JQ0894; JQ0894.
InterPro: IPR003439; ABC_transportr.
InterPro: IPR005289; GTP-binding_dom.
InterPro: IPR003405; SMC_C.
InterPro: IPR003395; SMC_N.
Pfam: PF02453; SMC_N; 1.
Pfam: PF02483; SMC_C; 1.
TIGREMS: TIGR00650; MG442; 2.
ATP-binding: Coiled coil.
NP_BIND 32 39 ATP (POTENTIAL).
FT DOMAIN 169 224 COILED COIL (POTENTIAL).
FT DOMAIN 231 400 COILED COIL (POTENTIAL).
FT DOMAIN 569 821 COILED COIL (POTENTIAL).
FT DOMAIN 884 912 ALA/ASP-RICH (DA-BOX).
SQ SEQUENCE 979 AA; 110566 MW; 30D51C56B56280F4 CRC64;

Query Match 22.5%; Score 85; DB 1; Length 979;
Best Local Similarity 30.9%; Pred. No. 4.2;
Matches 21; Conservative 15; Mismatches 22; Indels 10; Gaps 2;

QY 17 DLKNKVAOLKKVRSL-----KDKAAELKQEVSRLENEIEDLKAKIGDLNNTSGI--- 66
    | : : : : | : : : : | : : : : | : : : : | : : : : |
DB 323 DQYKTEIEIKKQVESLKIQTNASKQREIeldQQTRLNKANSUKLQENDINKEIGVLE 382

QY 67 RRPAAKLN 74

```





```
CC CC -1- FUNCTION: Required for the ATP- or CTP-dependent degradation of
CC CC proteins, but not small peptides, by the 20S proteasome (By
CC CC similarity).
CC CC -1- SUBUNIT: Homoheptamer (Potential).
CC CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC CC -1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR EMBL; AP000063; BAA81022.1; -.
DR InterPro; IPR003593; AAA_ATPase.
DR Pfam; PF00004; AAA; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00674; AAA; FALSE_NEG.
DR TIGRFAMs; TIGR01242; 26Sp45; 1.
KW Proteasome; ATP-binding; Complete proteome.
FT NP_BIND 180 187 ATP (POTENTIAL).
SQ SEQUENCE 409 AA; 44683 MW; 72F1274F0AB885DD CRC64;

Query Match 21.58; Score 81; DB 1; Length 409;
Best Local Similarity 37.38; Pred. No. 3.7;
Matches 19; Conservative 11; Mismatches 19; Indels 2; Gaps 1;

QY 3 GSHHHHGGSMASGLKNKVAQLKRVSLKDKAAELKQEVSRLENEIEDL 53
DB 8 GSRSHRHGGHSERDVEIRI--LADKVRSLTKELISLQKELEYKNEITKL 56

RESULT 12
Y334_METJA
ID Y334_METJA STANDARD; PRT; 102 AA.
AC Q57780;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein M30334.
GN M30334.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC -----
DR EMBL; U67487; AAB98322.1; -.

-1- FUNCTION: Required for the ATP- or CTP-dependent degradation of
proteins, but not small peptides, by the 20S proteasome (By
similarity).
-1- SUBUNIT: Homoheptamer (Potential).
-1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
-1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AP000063; BAA81022.1; -.
InterPro; IPR003593; AAA_ATPase.
Pfam; PF00004; AAA; 1.
SMART; SM00382; AAA; 1.
PROSITE; PS00674; AAA; FALSE_NEG.
TIGRFAMs; TIGR01242; 26Sp45; 1.
KW Proteasome; ATP-binding; Complete proteome.
FT NP_BIND 180 187 ATP (POTENTIAL).
SQ SEQUENCE 409 AA; 44683 MW; 72F1274F0AB885DD CRC64;

Query Match 21.58; Score 81; DB 1; Length 409;
Best Local Similarity 37.38; Pred. No. 3.7;
Matches 19; Conservative 11; Mismatches 19; Indels 2; Gaps 1;

QY 3 GSHHHHGGSMASGLKNKVAQLKRVSLKDKAAELKQEVSRLENEIEDL 53
DB 8 GSRSHRHGGHSERDVEIRI--LADKVRSLTKELISLQKELEYKNEITKL 56

RESULT 12
Y334_METJA
ID Y334_METJA STANDARD; PRT; 102 AA.
AC Q57780;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein M30334.
GN M30334.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC -----
DR EMBL; U67487; AAB98322.1; -.

-1- FUNCTION: Required for the ATP- or CTP-dependent degradation of
proteins, but not small peptides, by the 20S proteasome (By
similarity).
-1- SUBUNIT: Homoheptamer (Potential).
-1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
-1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AP000063; BAA81022.1; -.
InterPro; IPR003593; AAA_ATPase.
Pfam; PF00004; AAA; 1.
SMART; SM00382; AAA; 1.
PROSITE; PS00674; AAA; FALSE_NEG.
TIGRFAMs; TIGR01242; 26Sp45; 1.
KW Proteasome; ATP-binding; Complete proteome.
FT NP_BIND 180 187 ATP (POTENTIAL).
SQ SEQUENCE 409 AA; 44683 MW; 72F1274F0AB885DD CRC64;

Query Match 21.28; Score 80; DB 1; Length 102;
Best Local Similarity 40.08; Pred. No. 1.1;
Matches 22; Conservative 8; Mismatches 17; Indels 8; Gaps 1;

QY 19 KKKVAQLKRVSR-----SLKDKAAELKQEVSRLENEIEDLKKIKGLDNLNTSG 65
DB 37 KKKVAQLKRVSR-----SLKDKAAELKQEVSRLENEIEDLKKIKGLDNLNTSG 91

RESULT 13
SMC1_YEAST
ID SMC1_YEAST STANDARD; PRT; 1225 AA.
AC P32908;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chromosome segregation protein SMC1 (DA-box protein SMC1).
GN SMC1 OR CHL10 OR YFL008W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94103320; PubMed=8276886;
RA Strunnikov A.V., Lariou V.L., Koshland D.;
RT "SMC1: an essential yeast-gene encoding a putative head-rod-tail
RT protein is required for nuclear division and defines a new ubiquitous
RT protein family.";
RL J. Cell Biol. 123:1635-1648(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=95400292; PubMed=7670463;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae.";
RL Nat. Genet. 10:261-268(1995).
CC -1- FUNCTION: INVOLVED IN CHROMOSOME SEGREGATION IN MITOSIS. COULD BE
CC PART OF A CHROMOSOME CONDENSATION MOTOR.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH SMC2 OR OLIGOMERS.
CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC -1- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
CC FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC -----
DR EMBL; L00602; AAA16595.1; -.
DR EMBL; D50617; BAA09230.1; -.
DR PIR; A49464; A49464.
DR PIR; S41804; S41804.
DR SGD; S0001886; SMC1.
DR InterPro; IPR003405; SMC_C.
DR InterPro; IPR003395; SMC_N.
DR Pfam; PF02463; SMC_N; 1.
DR Pfam; PF02483; SMC_C; 1.
DR Mitosis; ATP-binding; Coiled coil; Nuclear protein.
```

```
Best Local Similarity   28.1%; Pred. No. 7;
Matches 18; Conservative    15; Mismatches      24; Indels       7; Gaps     1;
```

Qy 3 GSHHHGHGSMASGLKKNKA-----QLKRKYRSLKDKAAELKQFVSRLENEIDLKA 55  
| | | : ||||| : | : | : ||||| : | : | : ||||| : | : | : ||||| :  
Db 367 GFRRNSEGTQGADLSTQAIDQINGKLNRVTKEINXEFHQIEKFSEVEGRIODEK 426

Qy 56 KIGD 59  
: |  
Db 427 YVED 430

RESULT 15  
HEMA\_IATHO

ID HEMA\_IATHO STANDARD; PRT: 565 AA.  
AC A17000; Q84002; Q84003;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;  
GN Hemagglutinin HA2 chain].  
OS Influenza A virus (strain A/Equine/Tokyo/71).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
CC Influenza A viruses; Influenzavirus A.  
NCBI\_TaxID=11418;  
[1]  
SEQUENCE FROM N.A.  
MEDLINE=89204899; PubMed=2705299;  
RA Kawacka Y., Bean W.J., Webster R.G.;  
RL "Evolution of the hemagglutinin of equine H3 influenza viruses.";  
RT Virology 169:283-292(1989).

-I- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION.

-S- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HA1 AND HA2) LINKED BY A DISULFIDE BOND.

-I- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

\*\*\*\*\* THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to licensel@isb-sib.ch).

\*\*\*\*\*

DR EMBL: M24720; AAA3111.1; ALT\_SEQ.  
DR PR: C34064; HMIVS3.  
DR HSP: P03437; IHFM.  
DR InterPro: IPR001364; Hemagglutin.  
DR Pfam: PF00509; Hemagglutin; 1.  
DR PRINTS: PR00329; HEMAGGLUTN12.  
DR ProDom: PD000225; Hemagglutin; 1.  
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.  
FT SIGNAL 1 16  
FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.  
FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.  
FT CARBOHYD 23 23 N-LINKED (GLCNAC.. ) (POTENTIAL).  
FT CARBOHYD 37 37 N-LINKED (GLCNAC.. ) (POTENTIAL).  
FT CARBOHYD 53 53 N-LINKED (GLCNAC.. ) (POTENTIAL).  
FT CARBOHYD 68 68 N-LINKED (GLCNAC.. ) (POTENTIAL).  
FT CARBOHYD 78 78 N-LINKED (GLCNAC.. ) (POTENTIAL).  
FT CARBOHYD 180 180 N-LINKED (GLCNAC.. ) (POTENTIAL).  
FT CARBOHYD 300 300 N-LINKED (GLCNAC.. ) (POTENTIAL).  
FT CARBOHYD 498 498 N-LINKED (GLCNAC.. ) (POTENTIAL).  
SQ SEQUENCE 565 AA; 63580 MW; 84B7D4AD70629B7A CRC64;

Query Match 21.1%; Score 79.5; DB 1; Length 565;  
Best Local Similarity 28.1%; Pred. No. 7;  
Matches 18; Conservative 15; Mismatches 24; Indels 7; Gaps 1;

Qy 3 GSHHHGHGSMASGLKKNKA-----QLKRKYRSLKDKAAELKQFVSRLENEIDLKA 55  
| | | : ||||| : | : | : ||||| : | : | : ||||| : | : | : ||||| :

Db 367 GFRHONSEGTGAGDLKSTQAAIDQINGKLN RVIEKTNEKFHQIEKEFESEVEGRIQDLEK 426

Qy 56 KIGD 59

: |

Db 427 YVED 430

Search completed: December 18, 2002, 16:18:25  
Job time : 3.13584 secs

***This Page Blank (uspto)***

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	91	24.1	481	2	Q9LAX5	Q9Lax5 streptococ	
2	91	24.1	615	17	Q8TV66	Q8tv66 methanopyru	
3	89.5	23.7	356	13	Q98UK5	Q98uk5 brachydanio	
4	89.5	23.7	356	13	Q73679	Q73679 brachydanio	
5	89	23.6	315	13	Q98UK3	Q98uk3 brachydanio	
6	88.5	23.5	146	13	Q91891	Q91891 gallus gall	
7	88.5	23.5	286	13	Q57342	Q57342 coturnix co	
8	88	23.3	259	16	O66901	O66901 aquifex aeo	
9	87	23.1	324	13	Q9W6B1	Q9w6b1 brachydanio	
10	86	22.8	155	7	Q31405	Q31405 gallus gall	
11	86	22.8	609	17	Q8RXA4	Q8rxa4 methanopyru	
12	86	22.8	876	17	Q8RX14	Q8rx14 methanopyru	
13	85.5	22.7	286	13	Q42290	Q42290 gallus gall	
14	85.5	22.7	323	4	Q9H1F1	Q9h1f1 homo sapien	
15	85	22.5	321	7	Q31620	Q31620 gallus gall	
16	85	22.5	479	2	Q9LAX2	Q9Lax2 streptococ	

Q8TV66  
ID O8TV66  
PRELIMINARY:  
PRT: 615 AA.

```

AC O8TV66;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Predicted secreted protein.
GN MK1532.
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus.
OX NCBI_TaxID=2320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mezheva V.V., Makarova K.S., Polushin N.N.,
RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natarle D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Malykh A.G., Koonin E.V., Kozayavkin S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT and monophyly of archaeal methanogens."
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
DR EMBL; AF010443; AM02745.1; -.
KW Complete proteome.
SQ SEQUENCE 615 AA; 66941 MW; 77729702806BC922 CRC64;

Query Match 24.1%; Score 91; DB 17; Length 615;
Best Local Similarity 38.8%; Pred. No. 6;
Matches 19; Conservative 9; Mismatches 21; Indels 0; Gaps 0;

QY 15 SGDLKKNKVAQLKRVSLKDKAELKQEVSRLENEIEDLKAKIGDLNNT 63
Db 528 SNETOTTQLQNKVSTLBOGQELKQEVSKTSQEVSLKGSIDECKNT 576

RESULT 3
Q98UK5 PRELIMINARY; PRT; 356 AA.
AC Q98UK5;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Transcription factor MafB.
GN MAFB.
OS Brachydanio rerio (zebrafish) (zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21064923; PubMed=11134968;
RA Kajihara M., Kawauchi S., Kobayashi M., Ogino H., Takahashi S.,
RA Yasuda K.;
RT "Isolation, Characterization, and Expression Analysis of Zebrafish
RT Large MafB."
RL J. Biochem. 129:139-146(2001).
DR EMBL; AB006322; BAB21102.1; -.
DR InterPro; IPR004827; TF_bZIP.
DR InterPro; IPR004826; TF_Maf.
DR Pfam; PF03131; bZIP_Maf; 1.
DR SMART; SM00338; BRLZ; 1.
SQ SEQUENCE 356 AA; 40233 MW; DE4C96B62C058865 CRC64;

Query Match 23.7%; Score 89.5; DB 13; Length 356;
Best Local Similarity 22.2%; Pred. No. 4.7;
Matches 28; Conservative 13; Mismatches 22; Indels 63; Gaps 3;

QY 5 HHHHHH-----GSM-----ASGDLKKNKVAQLKRR 28
Db 215 HHHHHHPHGGQOHHGGVGGGLNVEDRFSDDQLVTMSVRELNLRLRGFTKDEVIRLKQK 274
QY 29 VRSLLKDK-----AAELKQEVSRLENEIEDLKAKIGDLN 61
Db 275 RRTLKNRGYAQSCRFRKRVQOKHLENEKTQLINQVQLKQEIINRLARERDAYKLKCEKLT 334

RESULT 4
Q98UK3 PRELIMINARY; PRT; 356 AA.
AC Q98UK3;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Transcription factor Val.
GN VAL OR VALENTINO.
OS Brachydanio rerio (zebrafish) (zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=98165393; PubMed=9425134;
RA Moens C.B., Cordes S.P., Giorgianni M.W., Barsh G.S., Kimmel C.B.;
RT "Equivalence in the genetic control of hindbrain segmentation in fish
RT and mouse."
RL Development 125:381-391(1998).
DR EMBL; AF006641; RAC18821.1; -.
DR ZFIN; ZDB-GENE-980526-515; val.
DR InterPro; IPR004827; TF_bZIP.
DR InterPro; IPR004826; TF_Maf.
DR Pfam; PF03131; bZIP_Maf; 1.
DR SMART; SM00338; BRLZ; 1.
SQ SEQUENCE 356 AA; 40243 MW; 07420DB0F6CD08F1 CRC64;

Query Match 23.7%; Score 89.5; DB 13; Length 356;
Best Local Similarity 22.2%; Pred. No. 4.7;
Matches 28; Conservative 13; Mismatches 22; Indels 63; Gaps 3;

QY 5 HHHHHH-----GSM-----ASGDLKKNKVAQLKRR 28
Db 215 HHHHHHPHGGQOHHGGVGGGLNVEDRFSDDQLVTMSVRELNLRLRGFTKDEVIRLKQK 274
QY 29 VRSLLKDK-----AAELKQEVSRLENEIEDLKAKIGDLN 61
Db 275 RRTLKNRGYAQSCRFRKRVQOKHLENEKTQLINQVQLKQEIINRLARERDAYKLKCEKLT 334

RESULT 5
Q98UK3 PRELIMINARY; PRT; 315 AA.
AC Q98UK3;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Smafl.
OS Brachydanio rerio (zebrafish) (zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21064923; PubMed=11134968;
RA Kajihara M., Kawauchi S., Kobayashi M., Ogino H., Takahashi S.,
RA Yasuda K.;
RT "Isolation, Characterization, and Expression Analysis of Zebrafish
RT Large MafB."

```







```
RA Malykh A.G., Koonin E.V., Kozhavkin S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RL and monophyly of archaeal methanogens.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
DR EMBL; AE010362; AA01904.1; -.
KW Complete proteome.
SQ SEQUENCE 876 AA; 103706 MW; D651E682271FD383 CRC64;

Query Match 22.8%; Score 86; DB 17; Length 876;
Best Local Similarity 45.5%; Pred. No. 23;
Matches 20; Conservative 8; Mismatches 12; Indels 4; Gaps 1;

Qy 17 DLKKNVAQLKRVSLKDKAAELKQEVSRLENEIEDLAKIGDL 60
    ||||| :||| | | : ||||| || | | : ||| : ||
Db 193 DLKSGKELKRVRELE-----ELKREKVELEPEVEELKRNEL 232

RESULT 13
O42290
ID O42290 PRELIMINARY; PRT; 286 AA.
AC O42290;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE bZIP transcription factor L-Maf.
GN L-MAF.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OC NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LENS;
RX MEDLINE=98192815; PubMed=9525857;
RA Ogino H., Yasuda K.;
RT "Induction of lens differentiation by activation of a bZIP
RT transcription factor, L-Maf.";
RL Science 280:115-118(1998).
DR EMBL; AF034570; AAC15781.1; -.
DR InterPro; IPR004827; TF_bZIP.
DR InterPro; IPR004826; TF_Maf.
DR Pfam; PF03131; bZIP_Maf; 1.
DR SMART; SM00338; BRUZ; 1.
DR SEQUENCE 286 AA; 32445 MW; 15DBD7B9B8DFA522 CRC64;

Query Match 22.7%; Score 85.5; DB 13; Length 286;
Best Local Similarity 25.9%; Pred. No. 8.1;
Matches 30; Conservative 12; Mismatches 25; Indels 49; Gaps 3;

Qy 4 SHHHHHG-----SMASGDL-----KNVAQLKRVSLKDK----- 35
    :||| ||| | | : ||| | | : |||
Db 157 AHHHHHHHLLERFSDQLVSNVSRVRLNQLRQFGSKVEVIRLKNRRTLNKRGYAQSC 216

Qy 36 -----AAELKQEVSRLENEIEDLAKIGDLNNTSGIRPA 70
    :||| ||| | | : ||| | | : |||
Db 217 RYKRVQRRHILENEKQLOQSQVEQLKQEVSLAKERDLYKEKYLEAARGFPREPS 272

RESULT 14
Q9H1F1
ID Q9H1F1 PRELIMINARY; PRT; 323 AA.
AC Q9H1F1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE DJ64411.1 (Kreissler (mouse) maf-related leucine zipper homolog).
GN KRM1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RA Ramsay H.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035665; CAB75863.1; -.
DR InterPro; IPR004827; TF_bZIP.
DR InterPro; IPR004826; TF_Maf.
DR Pfam; PF03131; bZIP_Maf; 1.
DR SMART; SM00338; BRUZ; 1.
DR SEQUENCE 323 AA; 35792 MW; A0F3C09F98936CB16 CRC64;

Query Match 22.7%; Score 85.5; DB 4; Length 323;
Best Local Similarity 23.1%; Pred. No. 9.2;
Matches 34; Conservative 9; Mismatches 19; Indels 85; Gaps 4;

Qy 5 HHHHHH-----GSMAS-----
    ||||| | | |
Db 162 HHHHHHQAAPPSSAASPAQQLPTSHPGPHATASATAAGNGSVEDRFSDQLVMSV 221

Qy 16 -----GDLKNVAQLKRVSLKDK-----AAELK 40
    | | | | :||| | | : ||| : ||
Db 222 RELNHLRGFTKDEVIRLQKRRTLKNRGYAQSCRYKRVQKKHLENEKTLQIQOVELK 281

Qy 41 QEVSRLENEIEDLAKIGDLNNTSGIR 67
    ||||| | | | | | | | | |
Db 282 QEVSLRERDAYKVKCEKLAN-SGFR 307

RESULT 15
Q31620
ID Q31620 PRELIMINARY; PRT; 321 AA.
AC Q31620;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE B-G protein (Fragment).
GN B-G.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OC NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91239571; PubMed=1903541;
RA Miller M.M., Goto R., Young S., Chirivella J., Hawke D., Miyada C.G.;
RT "Immunoglobulin variable-region-like domains of diverse sequence
RT within the major histocompatibility complex of the chicken.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:4377-4381(1991).
DR EMBL; M61862; AAA48627.1; -.
DR InterPro; IPR003600; IG_Like.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00410; IG_Like; 1.
DR NON_TER
FT 1
SQ SEQUENCE 321 AA; 36360 MW; 7F25BD29E39860C7 CRC64;

Query Match 22.5%; Score 85; DB 7; Length 321;
Best Local Similarity 26.8%; Pred. No. 10;
Matches 22; Conservative 18; Mismatches 24; Indels 18; Gaps 2;

Qy 7 HHHHGSMSGDLK-----NKVAQLKRVSLKDKAAELKQEVSRLENEIED 52
    | | | | :||| | | : ||| : |||
Db 185 HLAEKDLSTADLKLAAKLVEQREAVEERDSQLRQYKLGSRATNLKTLQKLENEIE 244

Qy 53 LKAKIGDLNNTSGIRPAKLN 74
    : | | | | | | | | | | | |
Db 245 VEKHLKKI-----GIRAPNKLH 262

Search completed: December 18, 2002, 16:21:03
Job time : 6.34459 secs
```

**THIS PAGE BLANK (USPTO)**